

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 7, 2002, 18:50:46 ; Search time 60.76 Seconds
(without alignments)
2332.623 Million cell updates/sec

Title: us-09-272-809-2

Perfect score: 6614
Sequence: 1 MNPNRSLIEDFLRVNFKR.....GRNQWLAYEGSQLPVHDEV 1276

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A.Geneset.032802:*

- 1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:*
- 2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:*
- 3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:*
- 4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:*
- 5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:*
- 6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:*
- 7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:*
- 8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:*
- 9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:*
- 10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:*
- 11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:*
- 12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:*
- 13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:*
- 14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:*
- 15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:*
- 16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:*
- 17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:*
- 18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:*
- 19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:*
- 20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:*
- 21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:*
- 22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6614	100.0	1276	21	AAAB26593
2	1003	15.2	212	21	AAAB26600
3	689	10.4	752	14	AAAR38153
4	661	10.0	740	14	AAAR38155
5	651.5	9.9	765	14	AAAR38150
6	635	9.6	340	12	AAAR14528
7	635	9.6	340	14	AAAR41019
8	635	9.6	340	16	AAAR5409
9	607	9.2	661	22	AAAG98885
10	447.5	6.8	574	14	AAAR38154
11	427	6.5	580	14	AAAR38151

12	375	5.7	844	21	AAAB26597
13	354.5	5.4	891	21	AAAB26596
14	351.5	5.3	1014	22	AAAG90902
15	349	5.3	481	21	AAAB26594
16	339.5	5.1	493	14	AAAR38156
17	332.5	5.0	950	21	AAAB26598
18	299.5	4.5	499	21	AAAB10471
19	299	4.5	387	22	ABG08893
20	299	4.5	387	22	ABG24364
21	297.5	4.5	498	20	AAAB9404
22	297.5	4.5	525	17	AAAR9466
23	295	4.5	532	14	AAAR42457
24	291	4.4	339	22	ABG06323
25	288	4.4	1371	21	AAAB26595
26	287	4.3	750	21	AAAB26599
27	284	4.3	411	22	ABG26062
28	280	4.2	895	22	ABG29771
29	273	4.1	509	22	ABG21574
30	267	4.0	745	22	ABG28573
31	265	4.0	371	21	AAAB16022
32	257.5	3.9	748	21	AAAB26592
33	250	3.8	1025	22	ABG29572
34	246	3.7	528	22	ABG15977
35	244.5	3.7	1142	19	AAAW50145
36	230.5	3.5	426	22	ABG30314
37	229	3.5	746	22	ABG10706
38	229	3.5	746	22	ABG22599
39	229	3.5	851	22	ABG25848
40	229	3.5	1404	22	ABG04193
41	228	3.4	736	22	ABG18312
42	224	3.4	296	21	AAAB15980
43	223.5	3.4	199	15	AAAR60986
44	218.5	3.3	1128	22	AAAG62367
45	216	3.3	1129	19	AAAW50144

ALIGNMENTS

RESULT 1	AAAB26593
ID	AAAB26593 standard; protein; 1276 AA.
XX	
AC	AAAB26593;
XX	
DT	01-FEB-2001 (first entry)
XX	
DE	Synechocystis sp phytochrome-related gene Cph2.
XX	
KW	Phytochrome; fluorescent apophytochrome; bilin; Cph.
XX	
OS	Synechocystis sp.
XX	
PN	WO2000056355-A1.
PD	28-SEP-2000.
XX	
PF	14-MAR-2000; 2000WO-US06607.
XX	
PR	19-MAR-1999; 99US-0272809.
XX	
PA	(RECC) UNITV CALIFORNIA.
XX	
PI	Lagaridis JC;
XX	
DR	WPI; 2000-602195/57.
XX	
PT	Fluorescent apophytochrome-bilin conjugates, e.g. Cph2 (derived from
PT	Synechocystis species) and phycoerythrobilin conjugate, useful as
PS	fluorescent markers for biological research -
XX	Example 1; Page 45-46; 52pp: English.

Synechocystis sp p
Synechocystis sp p
C glutamicum prote
Synechocystis sp p
Acetobacter diagan
Synechocystis sp p
Shewanella putrefa
Novel human diago
S. putrefaciens PK
Biosynthetic enzym
Enzyme involved in
Novel human diago
Synechocystis sp p
Synechocystis sp p
Novel human diago
Novel human diago
Novel human diago
Novel human diago
E. coli proliferat
Synechocystis sp p
Novel human diago
Novel human diago
Mesotetium caldar
Novel human diago
Novel human diago
Novel human diago
Novel human diago
Novel human diago
Novel human diago
Novel human diago
E. coli proliferat
A 22-26 kD antigen
Ap4 related amino
Oat phytochrome A

CC The present invention is related to fluorescent apophytochrome-bilin
CC conjugates, known as phytofluors. An apoprotein known as Cph2 from
CC *Synechocystis* species is used as the apophytochrome and the bilin is
CC preferably phycoerythrobilin. The phytofluors are useful as fluorescent
CC markers for biological research. The phytofluors have a long wavelength
CC absorption maxima, a high molar absorption coefficient and the
CC recombinant apoproteins can spontaneously assemble with a variety of
CC bilin chromophore precursors. The present sequence is a phytochrome
CC related protein from *Synechocystis* sp.

XX
Sequence 1276 AA;

Query Match 100.0%; Score 6614; DB 21; Length 1276;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1276; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MNPNRSLDFLRNNTKRRALTLRETLQVYEAKRIFGLGDRVAKIYFASGSGEVLAE 60
Db 1 mnpnrslfdlrnntkrraltlreqlqvyeaarkifglgdrvakikyfasgsgevlae 60
Qy 61 AVNRALPSLGLHPVPDIPPOAREELGNORKMIADVARRKKSHELSGRISTEHSN 120
Db 61 avnraalpsllglhfpvedippqareelgnrkmiadvarrkkshehsgristehsn 120
Qy 121 GHYTTVDSCHIQLYLLAMGSSLTVPVNOQDLGIMAVHHSKPRRTEOEMETWALLSK 180
Db 121 ghyttvdschqyllamgssltvpmqddqlgwimavhbskprfiefegewetallsk 180
Qy 181 EVSLAITQSLSROYHQOQVEALYQRLVETVAYQGDPRFTVOVALLEVGVGADEVGAVL 240
Db 181 evslaitqslsroyhqvgealvqrletvaygdprftvqvallevvgvgaevdgavl 240
Qy 241 YIAPDLTGSVAHQOMNRFDMGNMLETSLMOELMRGOPSAMPEMAVOSTWEKRPPT 300
Db 241 yiapdltgsvahqomnrfdmgnmletslmoelmrGOPsAMPEMAVOSTWEKRPPT 300
Qy 301 SVAPLPTNCVPHGYTGLGELBQRSDMIAPPSLSAENFOSFLIYPLAADQGVGSLILLR 360
Db 301 svaplptncvphgytlgelqrsdmiappslsaenfgsfllyplaadqgvgsllllr 360
Qy 361 KEKSLVKKMACKRGIDRNIIPLRSFEAMETOKIYPTWNSERKLAQVASTQLYMATIQ 420
Db 361 kekslvkkmckrgidrnllprlsfeawetqklyptwnserklaqvastqlymatiq 420
Qy 421 QFVRLITQCTAYDPLTOLPMNIIFFNRQLTALDALTDEGKMGVGLVAMDERKYNESF 480
Db 421 qfvrlitqctaydpltcqlpmvllfnrqltalldaldegkmvgvlylamdrfkynesf 480
Qy 481 GHKTGDLLOEVADRLNOKLSPLAASPLSRWHGDFITLLTQISDNQEMIPLCERLLS 540
Db 481 ghktgdllloevadrlnqklspilaasplsrwhgdftllltqisdngemiplcerlls 540
Qy 541 TFOEPFLQGGPIYLTASMGISTAPYDGETAESLKKFELILTRAKCOGKNTYOFRRPD 600
Db 541 tfgepfllqggpiyltasmgistapydgetaesllkfaeliltrakcgkntylfrrpd 600
Qy 601 SAPMLDRLTLESDDLROALTNOEFVLYPOVALDPTGKLGEVALVRMOHPRLGQVAPYV 660
Db 601 sapmldrltlesddlrgaltngelvylypqvaldptgkllgevalvrmohprlgvapyv 660
Qy 661 IPLAEELGLINHLGQWVLETACATHQHFRETGRRLRMANVISARQOFDEKMLNSVLECL 720
Db 661 iplaeelglinhlqgwvletacathqhfretgrllrmavlsarqfdekwlnsvlecl 720
Qy 721 KRTGMPEBDELEITESTLAMEDIKGVVLLHRLREGVQVAIDFGTGYSSLSILKQLPI 780
Db 721 krtgmpebdeleiteteslmedikgtvllhlreegvqvaiddgtgysslsilkqpi 780
Qy 781 HRLTIDSFVNDLNEGADTAIIQYVIDLANGLNEFPAEETIESBAQIQRLOKMGCHIQ 840
Db 781 hrltidsfvndlnegadtaiiqyvldlanglnlefpaeeetiesbaqiqrlokmgchiq 840

Qy 841 GFELTRPLPABAMTMYLYYQDILDFGPNPPLPKVALPETETEGAGNGVDRPLNSLRE 900
Db 841 gyltrplpaeamtmylyypqldfgrpplpkvalpeteteagngvdrplnslare 900
Qy 901 NPWTEKLDHYLLKRLDQNRNKEKLVKIANKRASLINDILYSTYVEYRQFLNDRV 960
Db 901 npwtekldhyllkerldqnrnkeklvikankiraslindilystyvevqfndrv 960
Qy 961 VLFKNSQMSQVWVESHNDCRSITINDEIDPCFKGHYLRLYRGRVAVSDIEKADLA 1020
Db 961 vlfknsqsgvveshndcrsitindeidpcfkghyrllyrgrvavsdiekadla 1020
Qy 1021 DCHKEELRHQYKANLVVYVFNENLWGLLAHECKPRRYQOEDQLIMELATQVALAI 1080
Db 1021 dchkeellrhqykanlvvynenlwgliaheckprryqeedqlimelatqvatal 1080
Qy 1081 HOGELXEQLEFANTRLOQISSLDALTYCGNRYLTFDSTLEREMORLQRTREPLALLCOVD 1140
Db 1081 hogelxeqlfetanrlqgissldaltvgnryllfstlerewqrtrreplallcdvd 1140
Qy 1141 FFKGFNDVNGHPADGRCLKTIADAMAKYAKRPTDLVARYGGEFAIILSETSLGAINVT 1200
Db 1141 ffkgrndvnghpagdrclkiadamakakrptdlvarggeefailsetslgainvt 1200
Qy 1201 EALQYEVANLAIPTVSGTGHVTSIGIAYVTPERHINPNAVLVAADLALYAKAKGNQ 1260
Db 1201 ealqyevanlaiphtvsgtghvtsigiaayvtperhinpnavlvaadlalyakakgnq 1260
Qy 1261 WLAYEGSOLPHVDGEV 1276
Db 1261 wlayegsqplhvddge 1276

RESULT 2

AAB26600
ID AAB26600 standard; peptide: 212 AA.
AC AAB26600;
XX 01-FEB-2001 (first entry)
XX

DE *Synechocystis* sp phytochrome-related gene Cph2-N197 peptide.
XX
KW phytochrome; fluorescent apophytochrome; bilin; Cph.
XX
OS *Synechocystis* sp.
XX
PN W0200056355-A1.
XX
XX 28-SEP-2000.
XX
PD 14-MAR-2000; 2000WO-US06607.
XX
PF 19-MAR-1999; 99US-0272809.
XX
PR (REGC) UNIV CALIFORNIA.
XX
PA
XX
PI Lagarias JC;
XX
XX WPI: 2000-602195/57.
XX
DR
XX

PT Fluorescent apophytochrome-bilin conjugates, e.g. Cph2 (derived from
PT *Synechocystis* species) and phycoerythrobilin conjugate, useful as
PT fluorescent markers for biological research -
XX

PS Claim 5; Fig 3; 52pp; English.

CC The present invention is related to fluorescent apophytochrome-bilin
CC conjugates, known as phytofluors. An apoprotein known as Cph2 from
CC *Synechocystis* species is used as the apophytochrome and the bilin is
CC preferably phycoerythrobilin. The phytofluors are useful as fluorescent
CC markers for biological research. The phytofluors have a long wavelength
CC absorption maxima, a high molar absorption coefficient and the

CC recombinant apoproteins can spontaneously assemble with a variety of
CC bilin chromophore precursors. The present sequence is a phylochrome
CC related peptide from *Synechocystis* sp.
SQ Sequence 212 AA;

Query Match 15.2%; Score 1003; DB 21; Length 212;
Best Local Similarity 100.0%; Pred. No. 3.1e-78;
Matches 197; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNPRSLDEPFRVNIKFRALTLRETLQVVEEARIFLGDVRYKTYFASDGSCEVIAE 60
DB 1 mnprrsledefrwnvknhraltlreclqvveearililgvdvayklykrasdgsgvrlae 60
QY 61 AVNRAALPSLLGLHPVEDIPPOAREELGNORKMIAVDVAHRRKKSHELSGRISPTESHN 120
DB 61 avnraalpsllglhfpvedipppareelngqrkmlavdvahrrkkshelesgrispteshn 120
QY 121 GHYTVDSCHIOYLAMGVLSLTVPVMDQDLGIMAVHNSKRRPRTEOEWEYMALISK 180
DB 121 ghytvdschilqyllamgvlsstlvpmqdgqlwylimavhnskprtrfegewetmalisk 180
QY 181 EVSLAITOSOLSROVHQ 197
DB 181 evslaitqsglsrqvqh 197

RESULT 3
ID AAR38153 standard; Protein; 752 AA.
AC AAR38153;

XX 13-OCT-1993 (first entry)

DE Acetobacter diuanylate phosphodiesterase PDEA2.

KW Cyclic diuanylate; diuanylate phosphodiesterase;

KM diuanylate cyclase; cellulose production; cdg2 operon.

XX Acetobacter xylinum.

OS Acetobacter xylinum.

XX Acetobacter xylinum.

XX Acetobacter xylinum.

XX Acetobacter xylinum.

XX Acetobacter xylinum.

XX Acetobacter xylinum.

XX Acetobacter xylinum.

XX Acetobacter xylinum.

XX Acetobacter xylinum.

XX Acetobacter xylinum.

XX Acetobacter xylinum.

XX Acetobacter xylinum.

XX Acetobacter xylinum.

XX Acetobacter xylinum.

XX Acetobacter xylinum.

PA (WEYE) WEYERHAEUSER CO.

XX Ben-Bassat A, Benziman M, Calhoon RD, Gelfand DH;

PI Tal R, Wong HC;

XX WPI: 1993-197062/24.

DR N-PSDB; AAQ43661.

XX Polynucleotide sequence from Acetobacter cdg operon - encodes

PT cyclic di:guanosine mono:phosphate degradation enzymes e.g.

PS 3-phosphodiesterase isozyme

XX Claim 5; Page 80-83; 98pp; English.

CC The amino acid sequence of protein PDEA2 was deduced from the 1st.

CC open reading frame of the cdg2 operon. The protein is a diuanylate

CC phosphodiesterase A, i.e. it enzymatically cleaves a single

CC phosphodiester bond in c-di-GMP to yield the linear dimer pGpG.

CC See also AAR38154.

XX Sequence 752 AA;

QY 19 HRAFLRETLQVVEEARIFLGVDRVKIKYKASDGSCEVIAEAVNRALPS-----L 70

DB 11 hgavgradvlefvtvdkpkl-----sadvllsamqadlpmvndehll 54

QY 71 LGHFPVEDIPPOAREELGNORKMIAVDVAHRRKKSHEL-----GRISPT-----E 117

DB 55 ifinaaeakiwgsreewygrnsvclpepdrndyinnretgyrlyvgsrewefr 114

QY 118 HSNCHY-----TVDSCHIOYLAMGVLSLTVPVMDQDLGIMAVHNSKRRPRFT 168

DB 115 rangeyvcgelstsrvgvndcklyla-----vmkd-----vteqsrqkll 157

QY 169 EOEWEYTMALSKREVSALATOSQSRQVHOQOEOALVORLETTVAOYGDREPTWOYALET 228

DB 158 vlqndvlgalasdsml-----gevgdlldcradsfvp----- 189

QY 229 VGOAVEADGAVLYTAPDITGSVAOHTQWNLKFRDMGNMLETSLMOELMGRQSAEMERMA 288

DB 190 -----gavgalmlldpsrrllsvs-----aspsmpkryaa 219

QY 289 VOSTWEKRRPFTSYA--PLPPTNCVPHGTYLGEJLQSRDWTAPPESTLAEMFQSLIVPL 346

DB 220 ldsmgqlrpeqlmtrampagsnlvwdya-----slarslglertccsllsr 268

QY 347 AADOOWGSLILRRKESLYKHMAGKRIGIDRRNLLPLSFPEWSETOQLVPTWNRSEKRL 406

DB 269 tgywmyfalyllrgdeadl-----awag----- 291

QY 407 AOVASTOL---YMAITQOFVTRLTQOTADPLQLPWWIIFNRQTLALDALY---EG 460

DB 292 -rvstmpfcalaieqsetrqlhlaqnsfslgl-----lntslmnlterllimggd 345

QY 461 KMGCVLVIAAMRFRKINSFGHKTGDDGLLOEVADRINKSLPLAAYSLSLRHWGDDGTI 520

DB 346 sqfslmvdldfrldlnahlvnadrlflaigrtrln---lvkddyivrsrggdehll 401

QY 521 LITQISDNQEMIPICERLSTFQEPFFLOGQPIVLTASMGISTAPYDEETSLKFAEI 580

DB 402 vrpccs-herckfaenlinalaiaarplygentliscvsgistfrangpdsesllshada 460

QY 581 ALTRAKCGKNTYQFRRPDSAPMILDRLLTLESDELQALTNQEFVLYQFOVYALDTGKLLG 640

DB 461 atrqakedgrglfrfagegkqvagdrilvlgalsrlsksmlnlngpyqvetmtg91y 520

QY 641 VEALVRMCHPRLGOVADRVFTRPLAEELGLNHLQOWLETCATNQHFRRTGRLKRAV 700

Db 521 vealgrwhprrlgnlpsrfflaaveetglaigrrwslaeacsqmrvkwdrgvrvptvav 580
Qy 701 NISAROFODEKWLNSVLECKRTGMPEDLELEITESLMMEDIKGTVLLHRLREGVQV 760
Db 581 nlsavhfrngrephianllkhghltptdlvleitesvmmssseeevlhrrldvyl 640
Qy 761 AIDDFGTGYSSLSLKQLPHRKIDKSFVNDLNEGADTAIIQYVIDLANGLNETVAE 820
Db 641 smdfgtgysisrlrlplrlteikidrsfindfndhnaavtmavlgisrlgmtvyle 700
Qy 821 GIESPAQIQRLOKMGCHLGCGYELTRPLP 849
Db 701 gvelegwrlleelncdvmgyltskplp 729

RESULT 4
ID AAR38155 standard; Protein: 740 AA.
AC AAR38155;
DT 13-OCT-1993 (first entry)
XX Acetobacter diiguanylate phosphodiesterase PDEA3.
XX Cyclic diiguanylate; diiguanylate phosphodiesterase;
KW diiguanylate cyclase; cellulose production; cdg3 operon.
XX Acetobacter xylinum.
XX
FH Key Location/Qualifiers
FT Misc-difference 119 /note= "Val deduced from GT"
FT Misc-difference 713 /note= "Trp deduced from GC"
FT Misc-difference 726 /note= "Ala deduced from CC"
XX
PN W09311244-A.
PD 10-JUN-1993.
XX
PF 14-OCT-1992; 92WO-US08756.
XX
PR 29-NOV-1991; 91US-0800218.
XX
PA (WEYE) MEYERHAUSER CO.
XX
PI Ben-Bassat A, Ben-Ziman M, Calhoon RD, Gelfand DH;
PI Tal R, Wong HC;
XX
DR WPI: 1993-197062/24.
DR N-PSDB; AAQ43662.
XX
XX Polynucleotide sequence from Acetobacter cdg operon - encodes
PT cyclic di:guanosine mono:phosphate degradation enzymes e.g.
PT 3-phosphodiesterase isozyme
XX
PS Claim 5; Page 86-88; 98pp; English.
XX
CC The amino acid sequence of protein PDEA3 was deduced from the 1st.
CC open reading frame of the cdg3 operon. The protein is a diiguanylate
CC phosphodiesterase A, i.e. it enzymatically cleaves a single
CC phosphodiester bond in c-di-GMP to yield the linear dimer pGpG.
CC See also AAR38156.
XX
SO Sequence 740 AA;

Query Match 10.0%; Score 661; DB 14; Length 740;
Best Local Similarity 26.5%; Pred. No. 1,1e-47;
Matches 208; Conservative 140; Mismatches 298; Indels 140; Gaps 20;

Qy 115 PTEHNGHYTTVD-----SCHIOYLLAMGVL---SSLTVPMQDQQLM--G 155
Db 59 pllhqdafatfeyersgshnrivgtstrevelftrsdgylcgelslskvnvndckrlffmg 118
Qy 156 IM-AVHHSKPRRTFOEMETALLSKESVLAITQSLSRQVHOQVQVDEALVQRIETVA 213
Db 119 ymkvntnesqgfk-----llllqndvlgqlasdmnl----qdvadllcrvestv- 164
Qy 214 QVGDREPTWQVALFTVGOAVEADGAVLYIAPLTGSVAOHYQWNLRFPMGWNLETSIMQE 273
Db 165 -----pgt-----vavlnltpdggqlrvls---spllkkyrtaeie---slyssselek 208
Qy 274 LMRGQPSAAMEPMAAVOSTWEKRPFTSVAPLPTNCVPHGYTGLCELDORSOWIAPESL 333
Db 209 l-----rvdpkhatriwmdysrlslglqgfcfcvstrsqgvk----- 249
Qy 334 SAENFOSFLIVPLADQQWGSLLILRREKSLVKHMKACKRGIDRRNIILPRLSFEAMETQ 393
Db 250 -----gflalysredqgrntwp-grivdscipfcaalafeg----- 283
Qy 394 KLVPTMNSERKLAQVASTQ-----LYMAITQOEVTBRLITQOTAYDPLTQLPNWIIFNRQ 448
Db 284 -----natqehishlanfdsltgllnrsvkhvlegmiskvgd-----nrq 324
Qy 449 LTLALLDALYEGKMGVGLVIAMDRFKRINESFGKTKGGLLOEVADRLNOKLSPLAATSP 508
Db 325 faifmld-----idfridindalghvayadqfllaelaarrs----lakedy 366
Qy 509 LLSRWHGSGFTILLQISDNQEMRPLCERLLSTFQEPFFLGQGPYLYLASMGISRPAVDG 568
Db 367 vlstsgdelfvvvvpd-cphkeatdfeahllasmtpmqllqntcltscslgltcypng 425
Qy 569 ETAEISLKFAEIALTRACOGKNTQYFRPDSAPMLRLTESDLROALTNOEFVLYEQ 628
Db 426 pdsesllstadvlqrakedgrgvfrfanlekngvagrlylgsalrslskgmnlhyq 485
Qy 629 POVALDTGKLLGVEALVWQHRPLGQVAPDVFIPLAEELGLINHLGQVLETTACATHOF 688
Db 486 pvyrtchtlslsgvealsrwhphrgnlfrsprflaaveetgqlaigrvslleacrgylkv 545
Qy 689 FRETGRRLMAVNISAROFODEKWLNSVLECKRTGMPEDLELEITESLMMEDIKGTVV 748
Db 546 drdgihpvtvavnlsvahfrnralpenlaalkhnlprslvleitesvmmssdridee 605
Qy 749 LHLRLREGVQVALIDFGTGYSSLSLKQLPHRKIDKSFVNDLNEGADTAIIQYVID 808
Db 606 vlgsirnlgsclsmddfgysslsrlrlplrlteikidrsfindfndhnaavtmavlg 665
Qy 809 LANGINLETVABGISEAQIQRLOKMGCHLGCGYELTRPLPAEAMMYLYTPQ-ILDFGP 867
Db 666 lgsrlgmtvvetegqrdlleklncdvmgylfakrplarddfekwmrwhqrlrgmlp 725
Qy 868 TPPLPK 873
Db 726 aapaak 731

RESULT 5
ID AAR38150 standard; Protein: 765 AA.
AC AAR38150;
DT 13-OCT-1993 (first entry)
XX Acetobacter diiguanylate phosphodiesterase PDEA1.
XX Cyclic diiguanylate; diiguanylate phosphodiesterase;
KW diiguanylate cyclase; cellulose production; cdg1 operon.
XX Acetobacter xylinum.
FH Key Location/Qualifiers

ID	AAG98865
ID	AAG98885 standard; Protein; 661 AA.
AC	AAG98885;
XX	
DT	26-SEP-2001 (first entry)
XX	
DE	E. coli growth and proliferation related protein sequence SEQ ID NO:355.
KW	Escherichia coli; growth; proliferation; microbial; antimicrobial; bacterial infection; microorganism.
OS	Escherichia coli.
PN	WO200134810-A2.
PD	17-MAY-2001.
PF	09-NOV-2000; 2000WO-US30950.
PR	09-NOV-1999; 99US-0164415.
PA	(ELIT-) ELITRA PHARM INC.
PI	Forsyth RA, Ohlsen K, Zyskind J;
DR	WPJ: 2001-335933/35.
DR	N-PSTDB; AAH84556.
PT	Novel nucleic acids that inhibit Escherichia coli proliferation, useful for screening for homologous genes and for designing expression vectors
PS	-
PP	Claim 19; Page 430-431; 522pp; English.
XX	
CC	AAH84373 to AAH84499 represent Escherichia coli growth and proliferation related DNA sequences (I). AAh84500 to AAh84670 encode the E. coli growth and proliferation related proteins given in AGS99078 and AAg98830 to AAg98999. (I) can be used as potential targets for the generation of new antimicrobial agents, and for identification of compounds which interact with the gene products of (I). In addition the expression of (I) and the purification of the proteins, the purified proteins can be used to generate reagents and screen small molecule libraries or other candidate compound libraries for compounds that can be further developed to yield novel antimicrobial compounds. In addition, nucleic acid probes complementary to (I) that are specific for particular species of microorganisms can be used to identify particular microorganism species in clinical specimens, therefore, providing a rapid and dependable method by which to identify the causative agents of a bacterial infection. Also, antibodies generated against proteins translated from mRNA transcribed under proliferation-regulated sequences can also be used to screen for specific microorganisms that produce such proteins in a species-specific manner. AAh84371 and AAh84670 represent sequencing primers used in the isolation of E. coli growth and proliferation related sequence, which are used in an example from the present Invention. Sequence 661 AA;
SQ	
Query Match	9.2%; Score 607; DB 22; Length 661;
Best Local Similarity	31.6%; Pred. No. 4.5e+43;
Matches 144; Conservative 110; Mismatches 165; Indels 36; Gaps 8;	
OY	401 RSERLAOVASPOLMAITQCFTFLILTYQTATVDPLTOLPMMIIIRNROLTLALLDALY-- 458
Dd	228 ragefirlrlant-----dsitglp-----nramqgllidmaih 261
OY	459 -EGKAVGVLVIMADRFRKRINESFGHKGTGDGLLEADVADRNOKLSPLAAYSPLTSRHHGGD 517
Dd	262 adnmkvygyvldlnlfrkvndayghlfqdqlrtvsial---lscl-ehdyatarpge 317
OY	518 FTILLITSNDOMEIPLCEERLLSTRFOEPEFLOGAPRIYLTAWSGISIAVPYGETAESLIKF 577

Db	318	flvlesntsqs-aleamaserlltrllprlgllievytscsvglalspshgsdstallrh	376
Oy	578	ALGLTFRACCGCKNNRYQAFRRPODASAPMLDRLTLESPLRQALTNQEFVLEFQOVALDTRK	637
Db	377	adltanytakeggrrgqfcvftremqgrvfyldtclnrlatendqdlvlnyqpkltw-7ge	435
Oy	638	ILGVEALVRMOPRLGOVAPADVETPLAEELGLINHLGQWVLETPACATHOHFRREGRRLR	697
Db	436	vrslaelvtrwpgperglldprldfdlisyaeagllvprgrvli-ldvrvqakwrtdkgnlr	494
Oy	698	MAVVISARQFODEKMLNSVLECLKRTGMPPEDELETESLMMDIKIGIVLLHRLREBG	757
Db	495	yavvnstargldqtlftalkvylqelnfeycpidvcltescllendlalsvlgqfsqil	554
Oy	758	VQVALDDFEGTSSLSILKOLPIHRLKIDKSFVNDLNGADTATIHQYVIDLANGINLET	817
Db	555	aqvhladdfytgysstlsqglarfrpidaikldqvtrvdthkqpvsgslvralavagaalnlgv	614
Oy	818	VAEGIESEAOLRLQKMGCHLQOQGYFLTRPRLPAEA	852
Db	615	laegyesakedaflcknglnerggflfakmpava	649
RESULT 10			
ID	AAK38154	standard: Protein; 574 AA.	
XX	AAK38154;		
AC			
XX	13-OCT-1993 (first entry)		
DT			
XX			
DE	Acetobacter diuanylate cyclase DGC2.		
XX			
KM	Cyclic diuanylate; diuanylate phosphodiesterase;		
KM	diuanylate cyclase; cellulose production; cdg2 operon.		
XX			
OS	Acetobacter xylinum.		
XX			
PN	MO9311244-A.		
XX	10-JUN-1993.		
PD			
XX	14-OCT-1992; 92WO-US08756.		
PF			
XX	29-NOV-1991; 91US-0800218.		
PR			
XX	(WEYE) WEYERHAEUSER CO.		
PA			
XX	Ben-Bassat A, Benziman M, Calhoon RD, Gelfand DH;		
PI	Tal R, Wong HC;		
PI			
XX	WPI. 1993-197062/24.		
DR	N-PSDB; AAQ43661.		
XX			
PT	Polynucleotide sequence from Acetobacter cdg operon - encodes		
XX	cyclic diuanosine mono:phosphate degradation enzymes e.g.		
PT	3-phosphodiesterase isozyme		
XX			
PS	Claim 5; Page 80-83; 98pp; English.		
XX			
CC	The amino acid sequence of protein DGC2 was deduced from the 2nd.		
CC	open reading frame of the cdg2 operon. The protein has diuanylate		
CC	cyclase activity, i.e. it enzymatically converts two molecules of		
CC	GTP to bis-(3'5')-cyclic diuanylic acid.		
CC	See also AAK38153.		
XX			
SQ	Sequence 574 AA;		

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 7, 2002, 18:52:41 ; Search time 25.57 Seconds
(without alignments)
1218.892 Million cell updates/sec

Title: US-09-272-809-2

Perfect score: 6614

Sequence: 1 MNPNSLEDPLRNVIKFKR.....GRNQWLAYEGSQLPHWGEV 1276

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA.*

1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep.*

2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep.*

3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep.*

4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep.*

5: /cgn2_6/ptodata/2/1aa/PCITUS.COMB.pep.*

6: /cgn2_6/ptodata/2/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	680	10.3	752	1	US-08-309-512-8
2	680	10.3	752	5	PCT-US92-08756A-8
3	662	10.0	740	1	US-08-309-512-10
4	662	10.0	740	5	PCT-US92-08756A-10
5	651.5	9.9	765	1	US-08-309-512-5
6	651.5	9.9	765	5	PCT-US92-08756A-5
7	450.5	6.8	574	1	US-08-309-512-9
8	450.5	6.8	574	5	PCT-US92-08756A-9
9	425	6.4	580	1	US-08-309-512-6
10	425	6.4	580	5	PCT-US92-08756A-6
11	339.5	5.1	493	1	US-08-309-512-11
12	339.5	5.1	493	5	PCT-US92-08756A-11
13	299.5	4.5	499	4	US-09-090-793-11
14	297.5	4.5	525	1	US-08-375-709-19
15	297.5	4.5	525	5	US-08-752-929-19
16	257.5	3.9	748	3	US-08-904-871-6
17	257.5	3.9	748	3	US-08-904-871-13
18	245	3.7	528	3	US-08-904-871-5
19	244.5	3.7	1142	3	US-08-904-871-12
20	225	3.4	554	3	US-08-904-871-1
21	222	3.4	611	3	US-08-904-871-4
22	216	3.3	1129	3	US-08-904-871-11
23	210.5	3.2	199	5	PCT-US94-02889-2
24	207	3.1	600	3	US-08-904-871-2
25	189	2.9	600	3	US-08-904-871-3
26	143	2.2	2154	2	US-08-841-349-4
27	133	2.0	1081	2	US-08-843-530B-18

28	133	2.0	1117	2	US-08-843-530B-33	Sequence 33, Appl
29	129	2.0	2860	2	US-08-826-267-2	Sequence 2, Appl
30	127.5	1.9	15281	2	US-08-471-119A-2	Sequence 2, Appl
31	125.5	1.9	656	3	US-08-738-000-4	Sequence 4, Appl
32	125.5	1.9	656	3	US-09-258-928-4	Sequence 4, Appl
33	125.5	1.9	660	3	US-08-738-000-2	Sequence 2, Appl
34	125.5	1.9	660	4	US-09-258-928-2	Sequence 2, Appl
35	125	1.9	3248	4	US-08-353-700-1	Sequence 1, Appl
36	125	1.9	3248	5	US-09-572-191-2	Sequence 1, Appl
37	122	1.8	1388	4	US-08-328-254-6	Sequence 2, Appl
38	121	1.8	2482	1	US-08-172-339-8	Sequence 6, Appl
39	118.5	1.8	870	4	US-08-874-678-33	Sequence 33, Appl
40	117.5	1.8	1362	2	US-08-643-839-33	Sequence 33, Appl
41	117.5	1.8	1362	3	US-07-736-178C-2	Sequence 2, Appl
42	117	1.8	655	1	US-09-199-637A-170	Sequence 170, App
43	116	1.8	1353	3	US-08-894-173-2	Sequence 2, Appl
44	116	1.8	1353	4	US-09-398-193-2	Sequence 2, Appl
45	116	1.8	1353	4	US-09-398-193-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-08-309-512-8
; Sequence 8, Application US/08309512
; Patent No. 5759828
; GENERAL INFORMATION:
; APPLICANT: Tal, Ronny
; APPLICANT: Benziman, Moshe
; APPLICANT: Gelfand, David H.
; APPLICANT: Ben-Bassat, Arle
; APPLICANT: Calhoon, Roger D.
; APPLICANT: Wong, Hing C.
; TITLE OF INVENTION: CYCLIC DIGUANYLATE METABOLIC ENZYMES
; NUMBER OF SEQUENCES: 63
; CORRESPONDENCE ADDRESS:
; ADDRESS: Pennie & Edmonds
; STREET: 2730 Sand Hill Road
; CITY: Menlo Park
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/309,512
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/800,218
; FILING DATE: 29-NOV-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Bortner, Scott R.
; REGISTRATION NUMBER: 34,298
; REFERENCE/DOCKET NUMBER: 8145-008
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 854-3660
; TELEFAX: (415) 854-3694
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 752 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYDROTHERMAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Acetobacter xylinum

QY 289 VOSTWEKPRPTSYA--PLPPTNCYPHGYTIGELBQRSDWIAPPSLSAENQSLIYPL 346
DB 220 LDSMQLPEQLEMLRANAGSNVYMDYA-----SLARSLERCCSSTIIR 268
QY 347 AADQWVSLILRKREKSLVHMAGKRGIDRRNIIPLRSFEAMETQKLVPTWNSERKL 406
DB 269 TGVWAGVATLIRGDEADL-----AMAO----- 291
QY 407 AQVASTOL--YMAITQOFTRLITQOTAYDPLTQLPNWIIFNRQTLALLDALY--EG 460
DB 292 -RVYSTSMFCALAEQSETRQIHALSNFDSLGL-----LNRSLNHTIERLIRMGD 345
QY 461 KMGVLYIAMRFRKINESFGHKTDGLQEVADRINKLSPLAAYSPLSRHWGDTI 520
DB 346 SQSLFMYDIDFRDINDALGHVADRFLIIGRIRH---LVKDDYIVSRSGDEPII 401
QY 521 LLTQISDQEMIPICERLSTFQEPFLOGOPILYLTASMGISTAYDETFESLTKFPEI 580
DB 402 VVPECS-HERAKRAENILNIAARPLQVGMTLSISCCVGISTFPANGPDSLSLHADA 460
QY 581 ALTRKCGKNTQOYFRRPDSAPMLDLTLLESDLRQALTNQEFVLXFOYALDTGKLLG 640
DB 461 ATROAKEDRGELFRPAGGEKQNOADRLVLSALRDSLKGLMLNLYQOYETMTGGLYG 520
QY 641 VEALVRWQHPRLGOVAPRVFTPLABELGLINHLGQVLETACATHQHFRETGRRLMAV 700
DB 521 VEALSRWHHPRLGNITPFRFAVAEETGOIEAIGWSLEECASQWVKVDRGVRPTAV 580
QY 701 NISAROFODEKMLNSVLECLRTGCMRPREDLELETESLMMEDIKGTAVLLHRLREGVQV 760
DB 581 NLSAVHRNRRLPREHIANLKLHNGITPRRLVEITTESYMMSSSTTEVLAIRLDYGL 640
QY 761 AIDFGTGYSSLILKOLPIHRLKIDKSFVNDLNEGADTAIOYVIDLANGLMEYAE 820
DB 641 SMDFGCGYSSLSLRTPLTEIKIDRSFINDFEHDNTAQAIVTMVIGISGLMTVYTE 700
QY 821 GIESEAOLORLQKMGCHGOGYFLTRPLP 849
DB 701 GVEEQOQRLLELHCDVMQGLFSKPLP 729

RESULT - 3
US-08-309-512-10
Sequence 10. Application US/08309512
Patent No. 5759828
GENERAL INFORMATION:
APPLICANT: Tai, Ronny
APPLICANT: Benziman, Moshe
APPLICANT: Gelfand, David H.
APPLICANT: Ben-Bassat, Arie
APPLICANT: Calhoun, Roger D.
APPLICANT: Wong, Hing C.
TITLE OF INVENTION: CYCLIC DIGUANYLATE METABOLIC ENZYMES
NUMBER OF SEQUENCES: 63
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 2730 Sand Hill Road
CITY: Menlo Park
STATE: California
COUNTRY: U.S.A.
ZIP: 94025
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/309,512
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/800,218
FILING DATE: 29-NOV-1991

ATTORNEY/AGENT INFORMATION:
NAME: Bortner, Scott R.
REGISTRATION NUMBER: 34,298
REFERENCE/DOCKET NUMBER: 8145-008
TELEPHONE: (415) 854-3660
TELEFAX: (415) 854-3694
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 740 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEetical: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Acetobacter xylinum
US-08-309-512-10

Query Match 10.0%; Score 662; DB 1; Length 740;
Best Local Similarity 26.3%; Pred. No. 4.3e-53;
Matches 207; Conservative 143; Mismatches 296; Indels 140; Gaps 20;

QY 115 PTEHNGHYTYVD-----SCHIOYLAMGVL---SSLTVVYVQDQOLM--G 155
DB 59 PTLHOAGHDAFVERSGSSHNRIYVGTSHREVEFTREDGEYICGLSLSKVYNDKRIFFMG 118
QY 156 IM-AVHHSKRPRFQEWEMTALLSKREVSALITQSLSRQVHOQOVQDALVQRIETVA 213
DB 119 VKNVNTNESQQRK-----ILLQNDVLAQALASDMI---QDVADLCRRVSEFV- 164
QY 214 QYGRPEWQVLALETQCAVEADGAVLYIADPLTGSVAQHOMNLFPMGMWLETSLMOE 273
DB 165 -FGT-----YAVMLITRPDQGLRVLS---SPTLKRRAISLE---SLYSSSELEK 208
QY 274 LMRGOPSAMEPMAVOSTWEKPRPTSVAPLPPTNCYPHGYTIGELBQRSDWIAPPSL 333
DB 209 L-----RVDPKNATRVWDSYRSL-----GISLG----- 232
QY 334 SAENQSFLLIYLAADQWVSLILRKESLIVKHMAGKRGIDRRNIIPLRSFEAMETQ 393
DB 233 -LQCFCTPVSSTRSQVKGIFALYRDEQGRNTPORIVDSCIPCALAFEO----- 283
QY 394 KLVPTWNSERKLAQVASTQ-----LYMAITQOFTRLITQOTAYDPLTQLPNWIIFNRQ 448
DB 284 -----NATQEHISHLANFDLITGLNRSVHKVIEGMSIQDS-----NRQ 324
QY 449 LTLALLDALYEGKMGVLYIAMDRKRNESFGHKTGDLQEVADRINKLSPLAAYSP 508
DB 325 FAIFMLD-----IDRFINDALGHVADQFLIEIARIRS-----TAKEDY 366
QY 509 LLSRWNGGFTIILTQISDQEMIPICERLSTFQEPFLOGOPILYLTASMGISTAYDE 568
DB 367 VLSRSGDEFEVYVVPD-CPRKEATDFAHLASMTMQIQNTLITLISGISITPYDPG 425
QY 569 ETAESLTKFAEIALTRAKCGKNTQOYFRRPDSAPMLDLTLLESDLRQALTNQEFVLXFO 628
DB 426 PDSLSLSTADVALRQAKEDRGVFRANLEKNOVADRLVLSALRDSLKGLMLNLYQOYETMT 485
QY 629 POVALDTGKLLQEVAVRWNQHPRLGQVAPRVFTPLABELGLINHLGQVLETACATHQHF 688
DB 486 FOVRHTLTLELGSVEALSRWHHPRLGNITPFRFAVAEETGOIEAIGWSLEECASQWVK 545
QY 689 FRETGRRLMAVNIASAROFODEKMLNSVLECLRTGCMRPREDLELETESLMMEDIKGT 748
DB 546 DDDGIHVPTVAVNLSAVHFRNRRLPREHTAALLKDHNLKPSRLTYEITTESYMMSSSTTE 605
QY 749 LILHRLREGVQVAIDDFGTGYSSLILKOLPIHRLKIDKSFVNDLNEGADTAIOYVID 808
DB 606 VLQSTRNIGCGLSMDDFGTGYSSLISRLRLPLTEIKIDRSFINDFEHDNTAQAIVTMVIG 665

QY 809 LANGINLETVAEGIESEAOLOKMGCHLGOGYFLTRPAPAEAMTYLYYPO-ILDFGP 867
Db 666 IGSRLGMYVTEGVEEQDRLKNCVMOGYLFAKPLAPDDEKMMRHQTIROMLP 725
QY 868 TPLPK 873
Db 726 AAPAK 731

RESULT 4

PCT-US92-08756A-10
Sequence 10, Application PC/TUS9208756A
GENERAL INFORMATION:
APPLICANT: Tal, Ronny
APPLICANT: Ben-Ziman, Moshe
APPLICANT: Gelfand, David H.
APPLICANT: Ben-Bassat, Arle
APPLICANT: Calhoon, Roger D.
APPLICANT: Wong, Hing C.
TITLE OF INVENTION: CYCLIC DIGUANYLATE
TITLE OF INVENTION: METABOLIC ENZYMES
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Limbach and Limbach
STREET: 2001 Ferry Building
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0,
SOFTWARE: Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/08756A
FILING DATE: 19921014
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/800,218
FILING DATE: 29-NOV-1991
ATTORNEY/AGENT INFORMATION:
NAME: Bortner, Scott R.
REGISTRATION NUMBER: 34,298
REFERENCE/DOCKET NUMBER: WEYR 20050 USA
TELEPHONE: 415-433-4150
TELEFAX: 415-433-8716
TELEX: 278356
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 740 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Acetobacter xylinum
PCT-US92-08756A-10

Query Match 10.0%; Score 662; DB 5; Length 740;

Best Local Similarity 26.3%; Pred. No. 4,3e-53; Indels 140; Gaps 20;

Matches 207; Conservative 143; Mismatches 236; Indels 140; Gaps 20;

QY 115 PTEHNGHYTTVD-----SCHIQYLLAMGYL---SLTPVPMQDOOLW--G 155
Db 59 PTLHAGHADAVERSRGSNNRIYVTSREVEFTSRSDGEYICGELSLKVVVNDKRIFFMG 118
QY 156 IM-AVHHSKRRTTEDEWEMTALLSKEVSLATIQSQLSRQVHOQOQVDEALVORLETTVA 213

Db 119 VMKNVTNESQORK-----ILLQNDVLOALASDMKI-----QDVADLLCRVESFV- 164
QY 214 QYGRPEFTMOYALETVGOAVEADGAVLYIAPDLTGSVAOHYOWNLRFPMGMNLETSLMOE 273
Db 165 -----PQT-----VAVMLILTPDGLRVLS---SPTLKKRYASLE---SLYSSSELEK 208
QY 274 LMRGOPSAMEBMAVOSTWEKPRPFTSVADPLPTNCVPHGYTLGELRQSDWIAPPESL 333
Db 209 L-----RVPKHTATRMWDSYRSL-----GISLG----- 232
QY 334 SAENFQSLIYPIADQMGWSLLILREKSLVKHMKRGIDRRNLIPLISFEAMETQ 393
Db 233 -----LQCCFCTPVSTRSGQVIGITALYSREDOGRTPORPLVDSCIPCALAFQ----- 283
QY 394 KLVPTRNSERKLAOVASTQ-----LYMAITQOFTRLITQOYVADPLTOLPMNITIFNRQ 448
Db 284 -----NATQEHISHLANFDSLTLGLNSSVHKYIEGMSIQDG-----NRQ 324
QY 449 LTLALLALYEGKMGVLIAMDRKRNIESFGHKTGGLLQEVADRLOKLSPLAAYSP 508
Db 325 FAIFMLD-----IDREINDALGHVYADQFLIEIARIRS---IAKEDY 366
QY 509 LLSRHGSGFTILLTQISDNQEMTPLCERLLSTQEPFLQGPYLYLASMISTAPYDG 568
Db 367 VLSRSGDEFFVWVPD-CPHKFATDFAEHLASMTMPQIQNTLTTSIGISTYPPNG 425
QY 569 ETASLTKFAELIATRAKCGKNTYQFYRPQDSAPMLDRLTLESPLROALTNQEFVLYQ 628
Db 426 PDSESLSTADVALROAKEDRGVFRFANLEKNVOARLVGSLRSLAKGMNLHLYQ 485
QY 629 POVALDTGKLLGVEALVWQHPRLGQVADVFIPLAELGLINHLGQVAVLETACATHQF 688
Db 486 PQRVHTLTLESGVEALSRHHPHGINFPFSRFIAVEETGOIEAIGRMSLEACRQIYKW 545
QY 689 FRETGRRLMAYNISARQODEKMLNSVLECKTKFGMPDELEITETPSLMMEDIKGVY 748
Db 546 DRDGIHPTVAVNLSAVHFRNRNALPETHAALLKDHNLPSLTYETTESVMMDSRDEE 605
QY 749 LHLRLREGVQVAIDFETGYSSILKQPLIRHLKIDKSFVNDLNGADTAIIOYVID 808
Db 606 VLOSIRNIGCGLSMDPDFGTGYSLSRLRLPLTEIKIDRSFINDEHTNQAQVATMAYIG 665
QY 809 LANGINLETVAEGIESEAOLOKMGCHLGOGYFLTRPAPAEAMTYLYYPO-ILDFGP 867
Db 666 IGSRLGMYVTEGVEEQDRLKNCVMOGYLFAKPLAPDDEKMMRHQTIROMLP 725
QY 868 TPLPK 873
Db 726 AAPAK 731

RESULT 5

US-08-309-512-5
Sequence 5, Application US/08309512
Patent No. 5759628

GENERAL INFORMATION:

APPLICANT: Tal, Ronny
APPLICANT: Ben-Ziman, Moshe
APPLICANT: Gelfand, David H.
APPLICANT: Ben-Bassat, Arle
APPLICANT: Calhoon, Roger D.
APPLICANT: Wong, Hing C.
TITLE OF INVENTION: CYCLIC DIGUANYLATE METABOLIC ENZYMES
NUMBER OF SEQUENCES: 63
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 2730 Sand Hill Road
CITY: Menlo Park
STATE: California
COUNTRY: U.S.A.
ZIP: 94025

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

```
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/309,512
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/800,218
FILING DATE: 29-NOV-1991
ATTORNEY/AGENT INFORMATION:
NAME: Bortner, Scott R.
REGISTRATION NUMBER: 34,298
REFERENCE/DOCKET NUMBER: 8145-008
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 854-3660
TELEFAX: (415) 854-3694
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 765 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Acetobacter xylinum
US-08-309-512-5
```

```
Query Match          9.9%: Score 651.5; DB 1; Length 765;
Best Local Similarity 33.3%: Pred. No. 4.5e-52;
Matches 161; Conservative 92; Mismatches 199; Indels 31; Gaps 7;
```

```
416 MAIQGFVTRITLTOQTAVDPLTOLPMNIFRNQLTLLDLYE--GKMGVLYAMDRF 473
283 LALBQHAKTKHTLQARVDSLTGLNRGALHR---VMDIIAOPGNFTLAIMDIDRF 338
474 KRINESFGHTGDLGLOEVADRLNOKLSPLAAYSPLSRMHGDTLLTQISDNQEMIP 533
339 RDINDALGHVYADQFLVLAIGRIRS---IAKDDVLSRSGDEFEVVVPP-CEGKQIEE 393
534 LCEELLSTFQEPFLOQPIYLTASMGISTAPYDGETAESLTKFAELTAKCOGKNTY 593
394 IAHKLETTIGRPDIQGNLTLSISCSIGSTFPANGPSESLSHADTAMROAKEDGRGIF 453
594 QFYRPDSAPMLDRLTLESDRALTNQEFVLYFQPOVALDTGKILGVEALVRMOHPRLG 653
454 RFANLEKNOVADRLVLSGALSRLDSLAGOMQLNHOPOVRTHTLLELSVEALSRMHHPRLG 513
654 QVADVPFPLAELGLINHLGQVNLLETACATNQHFFRETGRRLMAVNIISAROFODEKWL 713
514 NIFPSRFIAVAEETGOIEAIGRWSLLEACRQIVAKMDRGIHVPTVAVNLAVHFRNRALP 573
714 NSVLECLKRIGMPDEDELETESLAMEDIGTYVLLHRLREEGVQVADIDFGTGYSSL 773
574 EHAIALKLHMLKRSRTVETTESVMDNSRDLTEVLOSINNICGSLMDPFGTGYSSL 633
774 ILKOLPIHRLKIDKSFVNDLNEGADTAIIQYVIDLANGNLVEVAEIESEADLQRLQK 833
634 RLTLPLLEIKIDKSFINDFEYDTNQAQVIMAVIGISRLKMTYVTEGVEEQDRDLLEK 693
834 MCGHLGOGYFLTRPLPAEAMMTYLYPQILD---GTPPLPKVVALPETETEGAGCNV 889
694 LNCVMGGLYAKPLA-----PDLESWVRGGAFAVIR-----ELEAARAKG 737
890 DRP 892
738 GRP 740
```

RESULT 6

```
PCT-US92-08756A-5
Sequence 5, Application PC/TUS9208756A
GENERAL INFORMATION:
APPLICANT: Tal, Ronny
APPLICANT: Benzman, Moshe
APPLICANT: Gelfand, David H.
APPLICANT: Ben-Bassat, Arle
APPLICANT: Calhoon, Roger D.
APPLICANT: Wong, Hing C.
TITLE OF INVENTION: CYCLIC DISUGANYLATE
NUMBER OF INVENTION: METABOLIC ENZYMES
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESS: Limbach and Limbach
STREET: 2001 Ferry Building
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0,
SOFTWARE: Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/08756A
FILING DATE: 19921014
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/800,218
FILING DATE: 29-NOV-1991
ATTORNEY/AGENT INFORMATION:
NAME: Bortner, Scott R.
REGISTRATION NUMBER: 34,298
REFERENCE/DOCKET NUMBER: WEYR 20050 USA
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-433-4150
TELEFAX: 415-433-8716
TELEX: 278356
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 765 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Acetobacter xylinum
PCT-US92-08756A-5
```

```
Query Match          9.9%: Score 651.5; DB 5; Length 765;
Best Local Similarity 33.3%: Pred. No. 4.5e-52;
Matches 161; Conservative 92; Mismatches 199; Indels 31; Gaps 7;
```

```
416 MAIQGFVTRITLTOQTAVDPLTOLPMNIFRNQLTLLDLYE--GKMGVLYAMDRF 473
283 LALBQHAKTKHTLQARVDSLTGLNRGALHR---VMDIIAOPGNFTLAIMDIDRF 338
474 KRINESFGHTGDLGLOEVADRLNOKLSPLAAYSPLSRMHGDTLLTQISDNQEMIP 533
339 RDINDALGHVYADQFLVLAIGRIRS---IAKDDVLSRSGDEFEVVVPP-CEGKQIEE 393
534 LCEELLSTFQEPFLOQPIYLTASMGISTAPYDGETAESLTKFAELTAKCOGKNTY 593
394 IAHKLETTIGRPDIQGNLTLSISCSIGSTFPANGPSESLSHADTAMROAKEDGRGIF 453
594 QFYRPDSAPMLDRLTLESDRALTNQEFVLYFQPOVALDTGKILGVEALVRMOHPRLG 653
454 RFANLEKNOVADRLVLSGALSRLDSLAGOMQLNHOPOVRTHTLLELSVEALSRMHHPRLG 513
```


GENERAL INFORMATION:
APPLICANT: Calgene, LLC
TITLE OF INVENTION: Production of polyunsaturated fatty acids by expression
FILE REFERENCE: CGNE.131.01US
CURRENT APPLICATION NUMBER: US/09/090,793
CURRENT FILING DATE: 1998-06-04
EARLIER APPLICATION NUMBER: 60/048,650
EARLIER FILING DATE: 1997-06-04
NUMBER OF SEQ ID NOS: 66
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 11
LENGTH: 499
TYPE: PRT
ORGANISM: Shewanella putrefaciens
US-09-090-793-11

Query Match 4.5%; Score 299.5; DB 4; Length 499;
Best Local Similarity 29.0%; Pred. No. 3.1e-19;
Matches 101; Conservative 51; Mismatches 129; Indels 67; Gaps 11;

QY 933 KIRASLINDILYSTVTEVROFLNTD-----RVVLFKFNQSGGVVTESHNDFCRSIIN 987
DB 182 RIGANTKLNKV---TATSERLITNVGKPLKLVYHTNPPML-----DYSIIILL 232

QY 988 DEIDDPCEFGHYLRLY---REGGRAVSDIEKADLADCHKEELRHYOVKANLVPEVNE 1044
DB 233 VEMSLFLILAYFLYSFLVLPYR-KLASDIKMDKSRETKLRHYRPTTELKVAHTN- 290

QY 1045 NIMGLLIAHECKTPRYWOEDQLMELATOVVAIHOGELYEQUETANIRLQOISSIDA 1104
DB 291 -----ALM-----GTIOEQTKOLN---EGV-FIDK 311

QY 1105 LTQVNRRLFDSTLEREMORLORIREPLALLICDVDFKGFNDNGHPAGDRCLKKIDA 1164
DB 312 LTNIPNRRAFERLETYYQLARQOIGFTLLIADVDHREKYNDTGLHAGDALIKVAOT 371

QY 1165 MAKVAKRPDVLVARYGGEFATILSETSEGAIVNTEALQVEVNAIPIHTVSGT-GHVT 1223
DB 372 LSQGYRAEDICARFGGEFFIMLFRDIPDEPLQKLDAMLSFAELNLPNRSSTANYVT 431

QY 1224 LSGIAV-----YTPERHINPNALVKADLALYEAKAGRNQMLA 1263
DB 432 VSLGCVTVAVVDDEFKSESHIGSOALLIDKALYHAKACGRNALS 479

RESULT 14
US-08-375-709-19
Sequence 19, Application US/08375709
Patent No. 5683898
GENERAL INFORMATION:
APPLICANT: YAZAWA, Kazunaga
APPLICANT: YAMADA, Akiko
APPLICANT: KATO, Seishi
APPLICANT: KONDO, Kiyosi
TITLE OF INVENTION: Gene Coding For Elcosapentaenoic Acid
TITLE OF INVENTION: Synthesizing Enzymes and Process for Production of
TITLE OF INVENTION: Elcosapentaenoic Acid
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/375,709
FILING DATE: 20-JAN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/178,251
FILING DATE: 14-MAY-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 4-147945
FILING DATE: 15-MAY-1992
ATTORNEY/AGENT INFORMATION:
NAME: WEGNER, Harold C.
REGISTRATION NUMBER: 25,258
REFERENCE/DOCKET NUMBER: 53466/150/AAOK
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 525 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-375-709-19

Query Match 4.5%; Score 297.5; DB 1; Length 525;
Best Local Similarity 29.0%; Pred. No. 5.2e-19;
Matches 100; Conservative 50; Mismatches 128; Indels 67; Gaps 11;

QY 933 KIRASLINDILYSTVTEVROFLNTD-----RVVLFKFNQSGGVVTESHNDFCRSIIN 987
DB 208 RIGANTKLNKV---TATSERLITNVGKPLKLVYHTNPPML-----DYSIIILL 258

QY 988 DEIDDPCEFGHYLRLY---REGGRAVSDIEKADLADCHKEELRHYOVKANLVPEVNE 1044
DB 259 VEMSLFLILAYFLYSFLVLPYR-KLASDIKMDKSRETKLRHYRPTTELKVAHTN- 316

QY 1045 NIMGLLIAHECKTPRYWOEDQLMELATOVVAIHOGELYEQUETANIRLQOISSIDA 1104
DB 317 -----ALM-----GTIOEQTKOLN---EGV-FIDK 337

QY 1105 LTQVNRRLFDSTLEREMORLORIREPLALLICDVDFKGFNDNGHPAGDRCLKKIDA 1164
DB 338 LTNIPNRRAFERLETYYQLARQOIGFTLLIADVDHREKYNDTGLHAGDALIKVAOT 397

QY 1165 MAKVAKRPDVLVARYGGEFATILSETSEGAIVNTEALQVEVNAIPIHTVSGT-GHVT 1223
DB 398 LSQGYRAEDICARFGGEFFIMLFRDIPDEPLQKLDAMLSFAELNLPNRSSTANYVT 457

QY 1224 LSGIAV-----YTPERHINPNALVKADLALYEAKAGRNQ 1260
DB 458 VSLGCVTVAVVDDEFKSESHIGSOALLIDKALYHAKACGRNQ 502

RESULT 15
US-08-752-929-19
Sequence 19, Application US/08752929
Patent No. 5798259
GENERAL INFORMATION:
APPLICANT: YAZAWA, Kazunaga
APPLICANT: YAMADA, Akiko
APPLICANT: KATO, Seishi
APPLICANT: KONDO, Kiyosi
TITLE OF INVENTION: Gene Coding for Elcosapentaenoic Acid Synthesizing
TITLE OF INVENTION: Enzymes and Process for Production of Elcosapentaenoic
TITLE OF INVENTION: Acid
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.

COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/752,929
FILING DATE: 20-NOV-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/375,709
FILING DATE: 20-JAN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/178,251
FILING DATE: 14-MAY-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 4-147945
FILING DATE: 15-MAY-1992
ATTORNEY/AGENT INFORMATION:
NAME: WEGNER, Harold C.
REGISTRATION NUMBER: 25,258
REFERENCE/DOCKET NUMBER: 53466/150/AAOX
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 525 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-752-929-19

Query Match 4.5%; Score 297.5; DB 1; Length 525;
Best Local Similarity 29.0%; Pred. No. 5,2e-19;
Matches 100; Conservative 50; Mismatches 128; Indels 67; Gaps 11;

OY 933 KIRASINDILYSTVEYRQFLNLD---RVVLEKFNSSGSGOVVTESHNDFCRSIIN 987
DB 208 RIGANTKLNKV---TATSERLITNWDGKPLKLVLYHTNNOPPMV-----DYIITILL 258
OY 998 DEIDPCEKGYRLRY---REGRVAVSDIEKADLADCKELLRHYQVKANLVVPVENE 1044
DB 259 VEMSFLLILAYFLYSYFLVPRV-RLASDIKKMKSREIKLRHYPIVELVKVATHFN- 316
OY 1045 NIMGILLIAHECTPRYQWEDLQIMELATQVAIAIHQSELYEQLETTANIRLQOISSLDA 1104
DB 317 -----ALM-----GTIOETKOLN---EOV-FIDK 337
OY 1105 LTQGNRYLFDSTLREMRQRLQIREPLALLCDVDFEKGFMNDNGHPAGDRCLKRIADA 1164
DB 338 LTNINRRAFERLETTYCOLLRQOIGFTLLIADVDHEKEVNDYLGHLAGDEALIKVAQF 397
OY 1165 MAKVAKRPDLVARYGGEFFAILLETSEGAINTVETALQVEANLAIPHTVSGT-GHYT 1223
DB 398 LSGQFYRAEDICARFGGEFFIMLFRDIPDEPLQRIKIDAMLHSAELNLPHPNSSTANYVT 457
OY 1224 LSGIAV-----YTPRHINPALVKADLALEYEAKAKGRNQ 1260
DB 458 VSLGVCTVAVAVDDFEKFSSEHTIGSOAALIDAKALYHAKACGRNQ 502

Search completed: June 7, 2002, 18:55:44
Job time: 183 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 7, 2002, 18:53:06 ; Search time 37.51 Seconds
(without alignments)
3268.728 Million cell updates/sec

Title: US-09-272-809-2

Perfect score: 6614
Sequence: 1 MNPNSLEDFLRNVINKFHR.....GRNQWLAYEGSQLPHVDGEV 1276

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: PIR.71:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6614	100.0	1276	2 S75801	probable phytochro
2	967.5	14.6	880	2 AD1953	hypothetical prote
3	922	13.9	1415	2 C83070	conserved hypotet
4	882	13.3	748	2 D87632	sensory box/GGDEF
5	865	13.1	1244	2 S76102	hypothetical prote
6	862.5	13.0	1245	2 H83574	conserved hypotet
7	856.5	12.9	565	2 AB2334	hypothetical prote
8	830.5	12.6	760	2 E83610	conserved hypotet
9	821.5	12.4	696	2 S75626	hypothetical prote
10	819.5	12.4	788	2 AF0122	probable membrane
11	813.5	12.3	783	2 AC1834	hypothetical prote
12	811.5	12.3	783	2 B83232	conserved hypotet
13	811	12.3	685	2 H83428	conserved hypotet
14	810.5	12.3	687	2 A83019	conserved hypotet
15	809	12.2	687	2 F83100	conserved hypotet
16	801.5	12.1	840	2 S74707	nitrogen fixation
17	797	12.1	1578	2 S76238	hypothetical prote
18	780	11.8	611	2 AC2094	two-component resp
19	768	11.6	951	2 G82965	conserved hypotet
20	757	11.4	685	2 E82297	c-di-GMP phosphodi
21	752.5	11.4	856	2 E75292	GGDEF family prote
22	749.5	11.3	724	2 AH3133	GGDEF family prote
23	749.5	11.3	737	2 D98154	hypothetical prote
24	746	11.3	864	2 H83386	conserved hypotet
25	741	11.2	644	2 A96297	hypothetical 91.8K
26	739.5	11.2	644	2 AF2986	GGDEF family prote
27	737.5	11.2	732	2 S77565	hypothetical prote
28	737.5	11.2	821	2 AI2417	hypothetical prote
29	736	11.1	709	2 AE3468	di-guanilate cyclas

30	731.5	11.1	842	2 A87341	sensory box/GGDEF
31	729	11.0	772	2 G98327	hypothetical 91.8K
32	729	11.0	772	2 AF2955	GGDEF family prote
33	727.5	11.0	752	2 H98161	hypothetical 91.8K
34	727.5	11.0	779	2 AH3125	GGDEF family prote
35	717.5	10.8	564	2 E85895	hypothetical prote
36	714	10.8	742	2 H95270	probable kinase/es
37	714	10.8	1051	2 C95367	conserved hypotet
38	708	10.7	1021	2 AC2202	hypothetical prote
39	705.5	10.7	805	2 B75515	sensory box/GGDEF
40	703.5	10.6	749	2 A69861	conserved hypotet
41	700.5	10.6	809	2 B87260	sensory box/GGDEF
42	699.5	10.6	746	2 T35811	probable phosphodi
43	688.5	10.4	696	2 G97626	hypothetical prote
44	688.5	10.4	696	2 AI2849	GGDEF family prote
45	687	10.4	543	2 C84021	hypothetical prote

ALIGNMENTS

RESULT 1
S75801
probable phytochrome sl10821 - Synechocystis sp. (strain PCC 6803)
C:Species: Synechocystis sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
C:Accession: S75801
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, O. K.; Okumura, S.; Shimo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocys
s.
A:Reference number: S74322; MUID:97061201
A:Accession: S75801
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1276 <KAN>
A:Cross-references: EMBL:D64003; GB:AB001339; NID:g1001200; PIDN:BA10536.1; PID:d101
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Superfamily: phytochrome homology
C:Keywords: chromoprotein; phytochromobilin
F:1-416/Domain: phytochrome homology #status atypical <PHT>
F:129/Binding site: phytochromobilin (Cys) (covalent) #status predicted

Query Match	100.0%; Score 6614; DB 2; Length 1276;
Best Local Similarity	Pred. No. 0;
Matches 1276; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
Db	1 MNPNSLEDFLRNVINKFHRALTLRETLQVIVEARIFLGVDRVITYKRPASDGSCEVLAE 60
Qy	1 MNPNSLEDFLRNVINKFHRALTLRETLQVIVEARIFLGVDRVITYKRPASDGSCEVLAE 60
Db	1 MNPNSLEDFLRNVINKFHRALTLRETLQVIVEARIFLGVDRVITYKRPASDGSCEVLAE 60
Qy	61 AVNRALSLSLGLHPRVEDIPQAREELGNQKMTAVDVANHRKKSHELSGISTESHN 120
Db	61 AVNRALSLSLGLHPRVEDIPQAREELGNQKMTAVDVANHRKKSHELSGISTESHN 120
Qy	121 GHYTVVDSCHIOYLILAMGVLSLTVPVMDODOLMGIMAVHSHKPRFTQOEETALLSK 180
Db	121 GHYTVVDSCHIOYLILAMGVLSLTVPVMDODOLMGIMAVHSHKPRFTQOEETALLSK 180
Qy	181 EYSLAITGSQLSRQYHQOQVDEALVQRETTVAQYGDRETTQVLAETVGAVEADGAVL 240
Db	181 EYSLAITGSQLSRQYHQOQVDEALVQRETTVAQYGDRETTQVLAETVGAVEADGAVL 240
Qy	241 YIAPDLTSSVAGHQYQWNLRFPMGNLFTSLNQLMRGQPSAMEMMAVQSTWEXPRPFT 300
Db	241 YIAPDLTSSVAGHQYQWNLRFPMGNLFTSLNQLMRGQPSAMEMMAVQSTWEXPRPFT 300
Qy	301 SVAPLPPTNCVPHGYTLGELBORSDMIAPPELSAENFOSFLIVPLADQQVGSLLILR 360
Db	301 SVAPLPPTNCVPHGYTLGELBORSDMIAPPELSAENFOSFLIVPLADQQVGSLLILR 360

QY 361 KEKSLVHKWAKRGIDRRNIIPLRSEFAMEETOKLVPTWNSERKLAQVASTOLYMATIQ 420
D 361 KEKSLVHKWAKRGIDRRNIIPLRSEFAMEETOKLVPTWNSERKLAQVASTOLYMATIQ 420
QY 421 QFVTRLTQOYAYDPLTOLPMWITFNROLTALLDALYEGKMWGVLYVAMDFKINSF 480
D 421 QFVTRLTQOYAYDPLTOLPMWITFNROLTALLDALYEGKMWGVLYVAMDFKINSF 480
QY 481 GHKRGDGLQEVADRNLNOKISPLAAYSPILSRWHDGFTILLTOISDQEMPLICERILS 540
D 481 GHKRGDGLQEVADRNLNOKISPLAAYSPILSRWHDGFTILLTOISDQEMPLICERILS 540
QY 541 TFOEPEFLOGOPRYLTASMGISTAPYDGETAESLLKFAEIALTRAKCOGKNTYQFYRPD 600
D 541 TFOEPEFLOGOPRYLTASMGISTAPYDGETAESLLKFAEIALTRAKCOGKNTYQFYRPD 600
QY 601 SAPMLDRLTESDLRQALTNQEFVLYFQPOVALDQTKLLGYEALVRMOPRLGOVAPYF 660
D 601 SAPMLDRLTESDLRQALTNQEFVLYFQPOVALDQTKLLGYEALVRMOPRLGOVAPYF 660
QY 661 IPLAEELGLINHLGOWVLETACATQHOFRETGRRLRMAVNISARQFODEKMLNSVLECL 720
D 661 IPLAEELGLINHLGOWVLETACATQHOFRETGRRLRMAVNISARQFODEKMLNSVLECL 720
QY 721 KRTGMPEDLELETESIMEDIKGTVLLHRLREEGVOVALDQGTGYSLSILKQDPI 780
D 721 KRTGMPEDLELETESIMEDIKGTVLLHRLREEGVOVALDQGTGYSLSILKQDPI 780
QY 781 HRLTIDSFVNDLNEGADTAIOYVIDLANGLNLETABEGIESAOQORLOKMCCHIGQ 840
D 781 HRLTIDSFVNDLNEGADTAIOYVIDLANGLNLETABEGIESAOQORLOKMCCHIGQ 840
QY 841 GYFLTRPLPAPAMMYLYXPOILDFGPPPLPKVALPETETAGOGNDRPLPSLARE 900
D 841 GYFLTRPLPAPAMMYLYXPOILDFGPPPLPKVALPETETAGOGNDRPLPSLARE 900
QY 901 NPWTEKLDYVLLERLOQRNVKEKLYLKIANKIRASININDIIXSTVEYKQFLNDRV 960
D 901 NPWTEKLDYVLLERLOQRNVKEKLYLKIANKIRASININDIIXSTVEYKQFLNDRV 960
QY 961 VLFKFNQSGOVVTESHNDCRSITINDEIDPCFKGHYLRGRRAVSDIEKALA 1020
D 961 VLFKFNQSGOVVTESHNDCRSITINDEIDPCFKGHYLRGRRAVSDIEKALA 1020
QY 1021 DCHKEILRHQYOKANLYPVVFENELMGLLIAHECKTPRYQOEEDLOLMELATQVATAI 1080
D 1021 DCHKEILRHQYOKANLYPVVFENELMGLLIAHECKTPRYQOEEDLOLMELATQVATAI 1080
QY 1081 HOGELYEOLLETANIRLOOISSLDALTQVGNRYLFDSTLEREMORLQIREPLALLCDVD 1140
D 1081 HOGELYEOLLETANIRLOOISSLDALTQVGNRYLFDSTLEREMORLQIREPLALLCDVD 1140
QY 1141 FFKFENNYGHPADRCCKTADAMAKYAKRPTDVARVYGGEEPAIILSETSLGCAINVT 1200
D 1141 FFKFENNYGHPADRCCKTADAMAKYAKRPTDVARVYGGEEPAIILSETSLGCAINVT 1200
QY 1201 EALOVEVANLAIPTVSTGHTVLSIGIAYVTPERHINPNALVKAADLALYEAARAKGNQ 1260
D 1201 EALOVEVANLAIPTVSTGHTVLSIGIAYVTPERHINPNALVKAADLALYEAARAKGNQ 1260
QY 1261 WLAYEGSOLPHVDEGV 1276
D 1261 WLAYEGSOLPHVDEGV 1276

RESULT 2

AD1953
hypothetical protein a111175 [imported] - Anabaena sp. (strain PCC 7120)
C:Species: Anabaena sp.
A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002
C:Accession: AD1953

R:Kaneke, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriju
Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AD1953
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-880 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA873132.1; PID:917130522; GSPDB:GN00179
A:Experimental source: strain PCC 7120
A:Gene: a111175

Query Match 14.6%; Score 967.5; DB 2; Length 880;
Best Local Similarity 36.4%; Pred. No. 8.56-54;
Matches 248; Conservative 108; Mismatches 232; Indels 93; Gaps 16;

QY 188 QSQLSRQYHQOQVDEALVQRLTETVAOYGDREPFW-----QYALETGOAVAD 236
D 261 OKQAEELHOR-----DRLLQAVAEAN-----WLVEVMYDPAIEKALAVGGAQAD 309
QY 237 GAVLY--IAPDLTGSVAQHYQMNRFDM-GWLETSL--NQEIMRGOPSAMEPMAAYQS 291
D 310 RAYLFKNHHPHPTGKRAV---SLOFEWYGSDDLQPSIDQWQNU-----YQS 352
QY 292 TWEXRPPTSVAAPLPTPCVPHGYTLGELQORSDMIAPESISAEFQSPFLVPLAADOQ 351
D 353 T-ELSRKYSVLS-----SGKSINELRRKL-VAERELLENGIOSLLVPLGLEDO 401
QY 352 WYGSLLILRKESLVKHWAKRGIDRRNIIPLRSEFAMEETOKLVPTWNSERKLAQVAS 411
D 402 FMGYLGIA-----DCTQERY--WSRHEESTLITMA 429
QY 412 TOLYMATIQOYVTRLTQOYAYDPLTOLPMWITFNROLTALLDALYEGKMWGVLYVAMD 471
D 430 ASISGAMORQOYEEKIRQALHDLTLGLPNRLLENELAKALANASQESLAVWFLLD 489
QY 472 RFRKRSFHKGTGDLQEVADRNLNOKISPLAAYSPILSRWHDGFTILLTOISDQEM 531
D 490 RFKVINDLGHTLGDRLQSVQORLDCRS-----GDVYSRWGDEFTILLPHISYIEV 545
QY 532 IPLCERILSTQEPPEFLOGOPRYLTASMGISTAPYDGETAESLLKFAEIALTRAKCOGN 591
D 546 TQACARILKALEVEFHLHGHELYSASLGIALLDHNSPDALILKHADALYHAKDEERN 605
QY 592 TYQFYRPQDSAPMLDRLTESDLRQALTNQEFVLYFQPOVALDQTKLLGYEALVRMOPR 651
D 606 NYQFYTTSLSGKTPBELTLEKSLRYALEKQELKYQYQVRVNLITGQITGMALRMQHPRE 665
QY 652 LGOVAPDVFILAEELGLINHLGOWVLETACATQHOFRETGRRLRMAVNISARQFODE 710
D 666 MGIYAPSFIFLIAEETGILIPIGEMVLTACOMQR-AMQEAELRPYQAVANLSKQFQRP 724
QY 711 KWLNSVLECLKRTGMPEDLELETESIMEDIKGTVLLHRLREEGVOVALDQGTGYS 770
D 725 TLLLEVSAILAETGLAPLELETESITPAIEDIFTRSVRLNLOQMGVDAIDDFGTGHS 784
QY 771 SLTLKQPIHRLTIDNSFVNDLNEGADTAIOYVIDLANGLNLETABEGIESAOQOR 830
D 785 SLSRQDLPLNLKIDGFIKALTDSKVAHIVKAIYVTLGRSLDRLTBAEGVEKQEBLDF 844
QY 831 LQKMGCHLGOGYFLTRPLPAE 851
D 845 LKSTINCQDVQSGFLFKPLSAE 865

RESULT 3

C83070
conserved hypothetical protein PA4601 [imported] - Pseudomonas aeruginosa (strain PAO
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C:Accession: S76102
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
O, K.; Okumura, S.; Shimo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*
S.
A:Reference number: S74322; MUID:97061201
A:Accession: S76102
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1244 <KAN>
A:Cross-references: EMBL:D65999; GB:AB001339; NID:g1001396; PIDD:BA010080.1; PID:d101073
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match	13.1%;	Score 865;	DB 2;	length 1244;
Best Local Similarity	31.0%;	Pred. No. 5.9e-47;		
Matches 259;	Conservative 141;	Mismatches 286;	Indels 150;	Gaps 28;

```

0Y      145 VPVMOQ-----QIMIMAV-----HNSKPRR-----TEOBEMTAL-----177
Db      475 IPVADQTEETTPRLDSSMVAERKSPRIHQKGKALVSLRDITBOHARLAESEK   534
OY      178 -----LSKEVSIAITQSOLSRQVHOQ-----OVGEALVORLETTYVAQGRPETWQ   223
Db      535 KYRYIVELTSGIWILODDQQTTFRANQOLADMIGSVGEILEKNTATVLVIHHBPESQN   594
OY      224 YALETV-----GQAVEA-----DGAVLYIARDLTGSVAOHXOMNLRPMGNK--   265
Db      595 SHOKTIQSFPCVLPFNHQOYDVQFORRGDSVLM-----GLVSKSAMVD---OMGNRYG   645
OY      266 ---LETSLMOELMRGPSAMER-----MAAVOS-TWEKRPRTYSVALPRTNCVPNGYT   316
Db      646 ELAMLTDTTKKSAEOALSASEQRELGISIDQVWSADA VSFPTLYLANPTJAAVNGOS   705
OY      317 LGLEFORSD-W---IAPPESTSAENFOSFLIPLPAADOOWGSHLLIKRESLYKNHAGK   372
Db      706 LEVCQSQSNFMFEQYHNEDRILLEYHLQL---HEKDO-----TELEYRIYOPGKE   754
OY      373 RGIDRRNLPR-----LSFEAMEETOKLVPTNRSEKKLAQVASTOLYMAITOOFPYTRL   426
Db      755 RMLFRRSOLVRDGEHQPRIDS-----DITERKLA-----AEK   790
OY      427 ITQQTAVPYRLQRPWWIIENROLTLALLDALYECKM-VGVLIAMDREKRINESGHKTG   485
Db      791 LHYANNHDSLNLNRPBRFSFDRLGHALORNLRDRDLRFAPVLDLDGFKIINDSLGSCG   850
OY      486 DGLQEVADRINOKLSPIAAASPLISRWNHGDEFTILLPOISDNOMEPLRCERLSTEOPR   545
Db      851 DLLQOGIAHRQCIUR---EDTLARLGDEFTMLIENTICPREDEVIAVAQRINHQLQKP   906
OY      546 FFLOGQRYILTASMGIS-TARYDGETAESLKFAEIALTRAKCSQGKNTYOYRPRODSPA   604
Db      907 FNLGOEFETTSIGIALNHPRUGH-PODYLRADDTAYRAKKAAGKGYALTNOTMHNA   965
OY      605 LDRLTELROALTNENFYLFQPVNALDTGKLGVAVLRMOHPRLGGOVARVFTRPLA   664
Db      966 VOIRORENDLRALDRQLOLHYORIYCLTKGLOLQVAVLYRMGNHREGGILLPREFAIA   1025
OY      665 EELGLINHGWVLETA---CATHONEFRETYGRRLRAVNI SARODEKWLMSVLECL   720
Db      1026 EETGIIVMGMIIMEASROILLEKQSPOLS--HLQSVINWSROLDORLLKTVDIEIL   1083
OY      721 KRTEMPPDELEITESLAMEDIKGTIVULHRLABEGOVAIIDPGTGYSSLILKOLPI   780
Db      1084 SSTNAPADOLEKITESTLLDNLMLADVLSKLQRMNOIJSILDFTGYSLSYLHRRPI   1143
OY      781 HRLDKSFVNDLINEGADTAIIQYVIDLANGLINLEVATESEAOALORLQXXKHGLQ   840
Db      1144 NTFIVDSNFVTMEBNNONIAVHTIYLANHTGLGDVIAESIETERHLLTOHLHWIGCDQO   1203
OY      841 GYFLTRPLPARMMTYLYYPQILDGCPTRPLPKALBETEETEEAGQNVGDPRLPS   896

```

```

DB      1204 GYFFARPISDILDFL-----TIQIPRMSVPAV-TECH----GDKRKEEFS 1244

RESULT      6
H83574
conserved hypothetical protein PA0575 [imported] - Pseudomonas aeruginosa (strain PA0)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: H83574
R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;
adamson, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardly, K.; L
Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A:Reference number: A82950; MUID:20437337
A:Accession: H83574
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1245 <SMD>
A:Cross-References: GB:AE004493; GB:AE004091; NID:99946436; PIDN:AMG03964.1; GSPDB:GN
A:Experimental source: Strain PA01
C:Genetics:
A:Gene: PA0575

```

Query Match	13.0%;	Score 862.5;	DB 2;	Length 1245;
Best Local Similarity	29.1%;	Pred. No. 8.5e-47;		
Matches 263;	Conservative 154;	Mismatches 345;	Indels 141;	Gaps 24

```

0Y 20 RALLRETLQVIVEBARIELGVDRVKIKFAS-----DGSGEVLA-----59
   | | | | | : | : | : | | | | |
412 ROLLRL-GIMIDITEAKTELOAMRLSEQKFASFVHCPCDMVLVLANGADREFLAVNSTEEQ 470
0Y 60 -----EAVNNAALPSLLGHPVEDI-PQAREHIGQKMIADVAAHRRKXSHELSGR 112
   : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
471 QIGIADDAIGKTA--TELGI-WALPGLEPVLEKIGH-GNLWLEVEPLRRRNGSTFSAL 526
0Y 113 ISPTESHNGHYTV-----DSCHI--QYLLAM-----GYVLSLTVPMODQOL 153
   | | | | | : | : | : | | | | | : | : | : |
527 LS-AQHVALDQRPALVVVIRDTIHLVETQELLRISEKFNANAFHAPVPGDGLLSLEDGTL 585
0Y 154 WCIAMVHHSKPRRPFTEQEMETMALLSKVSLATIOSOLSRVHOQOVALQVLETYYA 213
586 ---IDVNGCFTRRLGVRDEVIISRSTLELGLMWPTEDRKRLISLVR-NHTLTOGFTAPVR 641
0Y 214 QYGDPEPMOYALETVQOAVADGAVLYIAPDLTGSVAQNHQMWLRFDMGNMLTSTMAOE 273
642 DRNGGIRQCEMSAHHI--SIDGEDCVLTIAARDIT-----EROLMOE 680
0Y 274 LMRGOPSAMERMA---AVOSTWEK-----PRPFTSVAPLPPTNCVPHGYTLGELQSDW 326
681 KIQ-OAAVVFVSTAGVWITDTRITAVNFAFSEIT-----GYS---EOBALG 725
0Y 327 IAPPELSAENQSFVLI---PLAADQOWGSLILREKSLVKNMGAKRGIDRRNLLPR 383
726 RSPSLLSGGHDSSFYLLMMNQLERDGHMOGEINNRRKTGSLYREW-----771
0Y 384 LSEFEAMEETQKLVPTW-----NRSEKRLAQVASTQLYMATIQOEVTRLLTIOQTAVDPLTQ 438
772 LTISAVHNPOGEITHFVGVFADISFLTKYQAQR-----LDYQAHNDRLTG 815
0Y 439 LPNWTIFNRQTLALDLALYEGKKWGVULVIMDBFKRINESFGHKTGGLLOEYADRLNQ 498
816 LPNRLLEFSRLNHALDEREBSRPGAVLFDLUDRKFKHINDSLGNPITGDLLIKAIAYELRD 875
0Y 499 KLSPLAASPLLSRWNGDGFITLLTIOISDNQEMRLPCEKRLSTEQEPPFLOGOPITYLAS 558
876 QLRDV-----DIVARLGGDEFTILLPLGHQESDAHVARKILLNAPTARFOADGHEFVSAS 931
0Y 559 MGISTAPYDGETAESLKFALFIALTRAKCOGKNUTYQFROPDSAPMLDRILTESDLQAL 618
932 VGIALFPPDQDAPLTVNADAAWYRAKSRBSRSFIEVYTRCLTVLAEARLLETELRRAL 991

```

Query Match	12.9%	Score 856.5	DB 2	Length 565
Best Local Similarity	40.4%	Pred. No. 5.4e-47		
Matches 193	Conservative 91	Mismatches 17	Indels 23	Gaps 4
OY	1PRLSFEAMEEFOKLVPMWRNSERKLAOVAASQYLMATIQGVTRLITQGFVADPDLTOLP	440		
Db	102 LPSLTFEYDEAGTITMADTSQTLHLKDE-----LIDYISNVDKLTGDP	144		
OY	441 NWIIFRQTLTLLDALYEGKNVGVLIAMDFRKRINESFGHKTGDSGLQEVADRNLQKLT	500		
Db	145 NRELLRERNQA-LSQYQSKMLAMVNIIDFTGISHALEHQSULLRAVAQRITSL	203		
OY	501 SPLAAYSPLSLSMWGDGFTLLITLTQISNOEMTPLCEBRLSTFOEPPELQGGPIYLTASMG	560		
Db	204 ----NOKDILMLSDSEFALAVEEIPSEISIKTLSQLTLITLNKPFPTIGNOIHHTASG	259		
OY	561 IS-TAPDGETAESLTKFAELITRAKQOGKTYQFYRRQDSAPMLDRTLESDDLQAL	619		
Db	260 ITINQPDMLHDVQLLOQAHVALYAKQGGRRQOHQFYSPEINAOERLALNELHGALE	319		
OY	620 NOEFLYFQPOVALDTGKLVEALVWRMCHPRLGOVAPVFTPLAEELGLINHLGOWYLE	679		
Db	320 RNEIIVVYQPIIDTQKOTAVEALVRMCHPRLGLVSPAKFTPIAEANGLIYEGWVLR	379		
OY	660 TACATGHQHFRETGRRLMAVNIISARQFODEKMWLSVLECLRTGMPDELEETESIM	739		
Db	380 TACIDNRTWQLAGLPIRKSVMLSARQFESNMLVELYSIIQESGLHPYLELVEVTESSL	439		
OY	740 MEDICGTIVLLHRLAEEGVQVAIDDFGTYSSSLSTIKOLPIHRLKIDKSFVNDLENGAD	799		

Query	Match	Similarity	12.6%	Score	830.5	DB	2	Length	760
Best Local Similarity	33.3%	Pred. No.	4.2e-45						
Matches	238	Conservative	104	Mismatches	248	Indels	125	Gaps	18
Qy	214	QYGDREPTMOVAL--ETFGVAQVEADGAVLYAPDLTGSVAOHYOMNLFEDGN--WLETS	269						
Db	120	ELGDPIETWKKRLHDEEYATVLEAHRNHL--QGLTLPDLHIYR--LHKRGDYRWLHS-	173						
Qy	270	LMQELMRQPSNAMEPMAAVOSTWEKRPRTSVAPLRPTNCVPYRGITLGEEO-----	322						
Db	174	-----RGR-----VLRDALGKPLHYGVA-----RDITLQRLKECHLRQAA	209						
Qy	323	-----RSDMI---APPESLASENQSF---LIV	344						
Db	210	AVFDSTREGVLTDAQAYIVHVNPSFERITGTRSDVGLKTRPALLRSRQDAAYQRLML	269						
Qy	345	PLADQQWVGSLLLRKESLWKHWAKRGIDRRNILEPRLSFEAMEETQKLVPTWNSER	404						
Db	270	ALREDDVMSGEIWNRRKSGEYLPQMLHTRAV--RNDGQLTNHYGVFSD--LSSIKRSN	325						
Qy	405	KLAQVASTQVLMATIQQVTRILTIQQRTAYDPVLQLPNNIIRNRQITLALLDALYEGKKVYG	464						
Db	326	ELDPLAH-----HDSLTGTPNNVLLRERIEQLKENGKDRYAGA	364						
Qy	465	VLVIAMDREKRRINESFGHKTDGDLQEVADRLLNOKLSPLAAYSPLLSKWHSDGFTILITQ	524						
Db	365	LLTLDLDFKHNNOSLGHITTDGMLKKEYSKRLQHQDLDRC---LLSLGGDEFALIV-E	419						
Qy	535	ISDNOEMTIPCEBRLISTQBPFFLOGOPITYLPASKNGISTADYDGETAESLIRAEIALTR	584						
Db	420	NDDPEAVARLLSORILLIDGNAPFELHICOPITYISASISGVSLEYPEDASDVHMLQHDADLFQ	479						
Qy	585	AKCGCKNRYOYRPQDSAPMLDLRLLESDLNQALTLNOEFVLYPORYALDQTKGLLGEVAL	644						
Db	480	AKDSGRNRYAFTTRYLIRARAHAVOESALRLHDELRVHYQVHDLASGRITGVESL	539						
Qy	645	VRMQHPRLGOVAPDVFIPLAEELGLINHLGOWVLEETACATQHOHFFERFGRRLR--MAVNIS	703						
Db	540	VRMQHPREGVLVPGEFVPAEBCGLIALDMWNVLRACRQKRE--WQNGVELEFVAVNVS	598						
Qy	704	ARQPDCKMINSVLECKRTGMPREDLELETESLAMDICTGVVLLHLRLAEBOVQAID	763						
Db	599	SRLLFRGGLEERIANALDESGLEPRYLLLEVYESVAMVDEFOSSLCLRLCLLIIVYNALD	658						

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 7, 2002, 18:54:01 ; Search time 19.14 Seconds

(without alignments)
2581.303 Million cell updates/sec

Title: US-09-272-809-2

Sequence: 1 MNPRLSEDFLRNRYNKFRH.....GRNOMLAVESQLPHVDGEV 1276

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6614	100.0	1276	PHY2_SYNY3	Q55434 synechocyst
2	821.5	12.4	696	Y195_SYNY3	P74101 synechocyst
3	708	10.7	827	Y41L_RHISN	P55552 rhizobium s
4	684.5	10.3	799	YDDU_ECOLI	P76129 escherichia
5	621	9.4	623	YD54_MYCTU	Q11024 mycobacteri
6	621	9.2	661	YCTR_ECOLI	P77334 escherichia
7	561	8.3	651	YHUK_ECOLI	P37649 escherichia
8	535	8.1	735	YNTC_AZOCA	Q04855 azorhizobi
9	400	6.0	528	YVCC_ECOLI	P33701 escherichia
10	386	5.8	782	YVIE_ECOLI	P75800 escherichia
11	370	5.6	747	YRGE_ECOLI	P77172 escherichia
12	368.5	5.6	307	YD57_MYCTU	Q11027 mycobacteri
13	366	5.5	516	YLAB_ECOLI	P74473 escherichia
14	361	5.5	507	YCGG_ECOLI	P75995 escherichia
15	358.5	5.4	532	YOAD_ECOLI	P76261 escherichia
16	356	5.4	518	RTN_ECOLI	P76446 escherichia
17	354.5	5.4	1000	Y041_SYNY3	O55445 synechocyst
18	323	4.9	362	YVHA_ECOLI	P21514 escherichia
19	296.5	4.5	753	BPHY_DEIRA	O91244 delinococcus
20	290	4.4	729	YFEA_ECOLI	P23842 escherichia
21	283.5	4.3	765	PHYA_ANASP	O91CC2 anabaena sp
22	280	4.2	452	YCDT_ECOLI	P75908 escherichia
23	273.5	4.1	751	PHYB_ANASP	O916X3 anabaena sp
24	265.5	4.0	1105	YRGE_ECOLI	P38097 escherichia
25	265	4.0	371	YALC_ECOLI	P21830 escherichia
26	262	4.0	1129	PHYB_SOLTU	P34094 solanum tub
27	261	3.9	1115	PHYE_PHANI	P55004 pharbitis n
28	258	3.9	1124	PHY_MOUSE	P33529 mougeotia s
29	257.5	3.9	748	PHY1_SYNY3	O55168 synechocyst
30	249.5	3.8	1136	PHY_PICAR	O40762 picea abies
31	247	3.7	1132	PHYB_TOBARC	P28130 nicotiana t
32	244	3.7	1124	PHYA_LATSA	P36573 lathyrus sa
33	242.5	3.7	728	BPHY_PSEAE	O9HW13 pseudomonas

34	237.5	3.6	1131	PHY_PINSY	Q41046 pinus sylve
35	237.5	3.6	1132	PHY1_PHPA	P36505 physcomitre
36	236.5	3.6	359	YHCK_BACSU	P54595 bacillus su
37	236	3.6	1124	PHYA_PRA	P15001 pisum sativ
38	235.5	3.6	1137	PHYC_ORYSA	O92W19 oryza sativ
39	234.5	3.5	1112	PHYE_ARATH	P42498 arabidopsi
40	233.5	3.5	1172	PHYB_ARATH	P14713 arabidopsi
41	231.5	3.5	1121	PHY2_CERPU	Q39557 ceratodon p
42	231	3.5	646	YHDA_ECOLI	P13518 escherichia
43	230.5	3.5	1128	PHYA_ORYSA	P10931 oryza sativ
44	229	3.5	1124	PHYA_CUCPE	P06592 cucurbita p
45	226	3.4	1118	PHY_ADICA	P42496 adiantum ca

ALIGNMENTS

RESULT	ID	PHY2_SYNY3	STANDARD:	PRT: 1276 AA.	
AC	Q55434;				
DT	16-OCT-2001 (Rel. 40, Created)				
DT	16-OCT-2001 (Rel. 40, Last sequence update)				
DT	01-MAR-2002 (Rel. 41, Last annotation update)				
DE	Phytochrome-like protein cph2 (Bacteriophytochrome cph2).				
GN	Cph2 OR SL0821.				
OS	Synechocystis sp. (strain PCC 6803).				
OC	Bacteria: Cyanobacteria; Chroococcales; Synechocystis.				
OX	NCBI_Taxid=1148;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=96127529; PubMed=8590279;				
RA	Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,				
RA	Sugiyama M., Tabata S.,				
RT	"Sequence analysis of the genome of the unicellular cyanobacterium				
RT	Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb				
RT	region from map positions 648 to 928 of the genome.";				
RL	DNA Res. 2:153-166(1995).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=97061201; PubMed=8905231;				
RA	Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,				
RA	Miyajima N., Hirosewa M., Sugiyama M., Sasamoto S., Kimura T.,				
RA	Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Nario K., Okumura S.,				
RA	Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,				
RA	Tabata S.;				
RT	"Sequence analysis of the genome of the unicellular cyanobacterium				
RT	Synechocystis sp. strain PCC6803. II. Sequence determination of the				
RT	entire genome and assignment of potential protein-coding regions.";				
RL	DNA Res. 3:109-136(1996).				
RN	[3]				
RP	CHROMOPHORE 1 BINDING, AND MUTAGENESIS OF CYS-129 AND HIS-130.				
RX	MEDLINE=20435261; PubMed=10978170;				
RA	Park C.-M., Kim J.-I., Yang S.-S., Kang J.-G., Kang J.-H., Shim J.-Y.,				
RA	Chung Y.-H., Park Y.-W., Song P.-S.;				
RT	"A second photochromic bacteriophytochrome from Synechocystis sp. PCC				
RT	6803: Spectral analysis and down-regulation by light.";				
RL	Biochemistry 39:10840-10847(2000).				
RN	[4]				
RP	IDENTIFICATION OF TWO BILIN LYASE DOMAINS.				
RX	MEDLINE=20510044; PubMed=11063585;				
RA	Wu S.-H., Lagarias J.C.;				
RT	"Defining the bilin lyase domain: lessons from the extended				
RT	phytyochrome superfamily.";				
RL	Biochemistry 39:13487-13495(2000).				
CC	"- FUNCTION: PHOTORECEPTOR WHICH EXISTS IN TWO FORMS THAT ARE				
CC	REVERSIBLY INTERCONVERTIBLE BY LIGHT: THE R FORM THAT ABSORBS				
CC	MAXIMALLY IN THE RED REGION OF THE SPECTRUM AND THE FR FORM THAT				
CC	ABSORBS MAXIMALLY IN THE FAR-RED REGION.				
CC	"- DOMAIN: Both bilin lyase domains bind with the bilin tetrapyrrole				
CC	chromophore precursor. The domain 1 shows red, far-red light				
CC	photo-reversibility. The domain 2 is photochemically inactive.				
CC	"- PFM: CONTAINS TWO COVALENTLY LINKED TETRAPYRROLE CHROMOPHORES.				

```
CC -1- SIMILARITY: BELONGS TO THE PHYTOCHROME FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D64003; BAA10536.1; -
DR Interpro: IPR001633; DUF2.
DR Interpro: IPR00160; DUF9.
DR Interpro: IPR003018; GAF.
DR Interpro: IPR001294; Phytochrome.
DR Pfam: PF00563; EAL; 1.
DR Pfam: PF00590; GAF; 3.
DR Pfam: PF00990; GGEF; 2.
DR SMART; SM00267; DUF1; 2.
DR SMART; SM00052; DUF2; 1.
DR SMART; SM00065; GAF; 3.
DR PROSITE; PS00245; PHYTOCHROME_1; FALSE_NEG.
DR PROSITE; PS50046; PHYTOCHROME_2; 2.
DR Photoreceptor; Phytochrome; Chromophore; Repeat; Complete proteome.
KW DCMAIN 1 197 BILIN LYASE DOMAIN 1.
FT DOMAIN 939 1075 BILIN LYASE DOMAIN 2.
FT BINDING 129 129 CHROMOPHORE 1 (PROBABLE).
FT BINDING 1022 1022 CHROMOPHORE 2 (POTENTIAL).
FT MOTAGEN 129 129 C-S: HOLOPROTEIN EXHIBITS NO
FT MOTAGEN 130 130 PHOTOCROMIC ACTIVITY.
FT MOTAGEN 130 130 H-S: CHROMOPHORE LIGATING ACTIVITY (IN
FT MOTAGEN 130 130 VITRO) IS 30-40% LOWER THAN WILD-TYPE.
FT MOTAGEN 130 130 H-S: CHROMOPHORE LIGATING ACTIVITY (IN
FT MOTAGEN 130 130 VITRO) IS ABOUT 10% MORE EFFICIENT THAN
FT MOTAGEN 130 130 WILD-TYPE.
SQ SEQUENCE 1276 AA; 144686 MW; 318CF3A73962D99E CRC64;
Query Match 100.0%; Score 6614; DB 1; Length 1276;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1276; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 481 GHKTGDGLQEVADRLNKLSPLAAYSPLSRMHGDGFTILLTQISDNQEMIPICERLLS 540
DB 481 GHKTGDGLQEVADRLNKLSPLAAYSPLSRMHGDGFTILLTQISDNQEMIPICERLLS 540
QY 541 TPQEPFPLQGPPIYTLASMGISTAPYDGETAESLKFEMIALTRAKCGKNTQGYRRQD 600
DB 541 TPQEPFPLQGPPIYTLASMGISTAPYDGETAESLKFEMIALTRAKCGKNTQGYRRQD 600
QY 541 TFOEPPFLOGQPIYTLASMGISTAPYDGETAESLKFEMIALTRAKCGKNTQGYRRQD 600
DB 541 TFOEPPFLOGQPIYTLASMGISTAPYDGETAESLKFEMIALTRAKCGKNTQGYRRQD 600
QY 601 SAPMLDRLTLESDDLROALTNOEFVLYPOVALDPTGKLGLVEALVRMHQPRILGQVAPVDF 660
DB 601 SAPMLDRLTLESDDLROALTNOEFVLYPOVALDPTGKLGLVEALVRMHQPRILGQVAPVDF 660
QY 661 IPLAEELGLINHLGQWVLETACATHQHFRETGRRLRNAVNI SARQFODEKWLNSVLECL 720
DB 661 IPLAEELGLINHLGQWVLETACATHQHFRETGRRLRNAVNI SARQFODEKWLNSVLECL 720
QY 721 KRTGMPPEDELLETESLAMEDIKGTVYLLRLREGQVVAIDPRTGYSSLSILKQPI 780
DB 721 KRTGMPPEDELLETESLAMEDIKGTVYLLRLREGQVVAIDPRTGYSSLSILKQPI 780
QY 781 HRLKIDKSFVNDLNEGADTAIIQYVIDLANGLNLETVAEGLIESEAQLQRLQKMGCHLQ 840
DB 781 HRLKIDKSFVNDLNEGADTAIIQYVIDLANGLNLETVAEGLIESEAQLQRLQKMGCHLQ 840
QY 841 GYFLTRPLPAEAMMTYLYYPOILDFGPTPLPKYALPETEDAGGNVGRPLPNSLRE 900
DB 841 GYFLTRPLPAEAMMTYLYYPOILDFGPTPLPKYALPETEDAGGNVGRPLPNSLRE 900
QY 901 NPWTEKLDHYLLKRLQORNVKELVLIANKIRASININDILYSTVEYRQFNTDRV 960
DB 901 NPWTEKLDHYLLKRLQORNVKELVLIANKIRASININDILYSTVEYRQFNTDRV 960
QY 961 VLEKFNQSGQVYTESHNDCRSIINDEIDDPCKGKGYRLRYRGRAVSDLEKADLA 1020
DB 961 VLEKFNQSGQVYTESHNDCRSIINDEIDDPCKGKGYRLRYRGRAVSDLEKADLA 1020
QY 1021 DCHKELNHYOVKANLVYVYVFNENKGLITAEKTRTYQOEEDLOLMELATQVALAI 1080
DB 1021 DCHKELNHYOVKANLVYVYVFNENKGLITAEKTRTYQOEEDLOLMELATQVALAI 1080
QY 1081 HOGELYOLETANIRLQOISSDALTOYGNRYLPFSTJEREMQRLQRIREPALLDCVD 1140
DB 1081 HOGELYOLETANIRLQOISSDALTOYGNRYLPFSTJEREMQRLQRIREPALLDCVD 1140
QY 1141 FFKGFNDYGHPRAGDRCLKIADAMAKVAKRPTDLVARYGGEFAI IISFSLGCAINVT 1200
DB 1141 FFKGFNDYGHPRAGDRCLKIADAMAKVAKRPTDLVARYGGEFAI IISFSLGCAINVT 1200
QY 1201 EALOYEVANLAIPIHTVSGTGHVTLISIGTAVYTPERHINPMLVYKADALALYEA KAKGRNO 1260
DB 1201 EALOYEVANLAIPIHTVSGTGHVTLISIGTAVYTPERHINPMLVYKADALALYEA KAKGRNO 1260
QY 1261 WLAYEGSOLPHVGEV 1276
DB 1261 WLAYEGSOLPHVGEV 1276
```

RESULT 2
Y195_SYNY3
ID Y195_SYNY3 STANDARD; PRT; 696 AA.
AC P7A101.
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 79.7 kDa protein s11895.
GN SLL1895.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxId=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;

RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hirosewa M., Sugiyama M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Shimp S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT *Synechocystis* sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
CC -1- SIMILARITY: BELONGS TO THE YEGE/YHDA/YHUK/YJCC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: D90912; BAA18187.1;
DR InterPro: IPR001633; DUF2.
DR InterPro: IPR000160; DUF9.
DR InterPro: IPR000253; FHA_domain.
DR InterPro: IPR000014; PAS.
DR Pfam: PF00563; EAL; 1.
DR Pfam: PF00498; FHA; 1.
DR Pfam: PF00990; GGDEF; 1.
DR SMART: SM00267; DUF1; 1.
DR SMART: SM00052; DUF2; 1.
DR SMART: SM00240; FHA; 1.
DR SMART: SM00091; PAS; 1.
DR PROSITE: PS50006; FHA_DOMAIN; 1.
DR Hypothetical protein; Complete proteome.
SQ SEQUENCE 696 AA; 79670 MW; 78EB2D40EA6801 CRC64;

Query Match 12.4%; Score 821.5; DB 1; Length 696;
Best Local Similarity 38.1%; Pred. No. 1.1e-45;
Matches 171; Conservative 89; Mismatches 184; Indels 5; Gaps 2;

407 AOVASTOLYMAITQOFVRLITQOTADPDLQPLPMIIFNQLTALLDALYEGKMGVL 466
239 SIFIRSYIFDITERKVIDQSIIHYQAFYDPLDPLRFLFKGLKGLSVNQNQSSVLGLV 298
467 VIANDREKRNESFGHKRGDGLDEVDRLNOKSLPLAAYSPLSRWHDGFTILLTOIS 526
295 LIGRELIQSLNDLGLSHVADVLKITER-----LSAHVRLLEDLCRMGDTFILLIQSCR 354
527 DNOEMIPLCERLSTFOEPFLQGOPIYLFTASMGISTAPYDQETASLLKFAEIALTRAK 586
355 NLDEIVEVRLRLAVLAKRPFIANNPFLQYAGIACIPNNGDNVEILLNRVGIALLNEVK 414
587 CGKNITQFYRPODSAPMLRLTLESDLQALNQEFVLYFPOVALYDTGKLLGVEALVR 646
415 DIGRQYCFEESMNSDHLERIOLEHALQALERDEFLLYOPTIDVQSGRLCGEALIR 474
647 WOHRLGOVADVPLPLAEELGLNHLGQWVLETRACATHOFPRFETGRRLMAVNISARQ 706
475 WOHRLHGVASGLFTGLTGLTIVGEMTMRFAFOHFMHMAVADDDFRIRAINLSPQ 534
707 FODEKWLNVDELCKRTGMPEDELEETTESIMMEDIKGVLTALHRLREEVOYAVIDFG 766
535 FQADDLPLTILRIALAESLPRHRELEETENIVQNVATONLINALQSHGIRLSMDFG 594
767 TGYSLSLTKOLPLHRLKIDKSPVNDLNEGADTAIIQYVIDLANGLLETVAIGIESEA 826
595 TGYSLSLTKLPFPNTKIDRSFTKIDILHTPKDAITIOAMLLNGFNLIANGIESEP 654
827 QLORLQMGCHLGGGYFLTRPLRAEMMT 855
655 QARCLTYDGLGKEMQGWFSHPL-SEAEIT 682

RESULT 3
YAIL_RHISN
ID YAIL_RHISN STANDARD; PRT: 827 AA.
AC P55552;
DT 01-NOV-1997 (Rel. 35. Created)
DT 01-NOV-1997 (Rel. 35. Last sequence update)
DT 16-OCT-2001 (Rel. 40. Last annotation update)
DE Hypothetical 91.8 kDa protein YAIL.
GN YAIL.
OS Rhizobium sp. (strain NGR234).
OC Plasmid sym pNGR234.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=394;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97305956; PubMed=9163424;
RA Freilberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,
RA Perret X.;
RT "Molecular basis of symbiosis between Rhizobium and legumes.";
RL Nature 387:394-401(1997).
CC -1- SIMILARITY: BELONGS TO THE YEGE/YHDA/YHUK/YJCC FAMILY.
CC -1- SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AE000083; AAB91764.1;
DR InterPro: IPR001633; DUF9.
DR InterPro: IPR000160; DUF9.
DR InterPro: IPR001610; PAC.
DR InterPro: IPR000700; PAS-assoc_C.
DR InterPro: IPR000014; PAS.
DR Pfam: PF00563; EAL; 1.
DR Pfam: PF00990; GGDEF; 1.
DR Pfam: PF00785; PAC; 2.
DR Pfam: PF00989; PAS; 3.
DR SMART: SM00267; DUF1; 1.
DR SMART: SM00052; DUF2; 1.
DR SMART: SM00086; PAC; 2.
DR SMART: SM00091; PAS; 3.
DR PROSITE: PS50113; PAS; 3.
DR PROSITE: PS50113; PAC; 2.
DR Hypothetical protein; Repeat; Plasmid.
KM DOMAIN 12 82 PAS 1.
FT DOMAIN 212 264 PAS 1.
FT DOMAIN 265 335 PAS 2.
FT DOMAIN 338 390 PAC 2.
SQ SEQUENCE 827 AA; 91818 MW; 0D6CBF1142027FAB CRC64;

Query Match 10.7%; Score 708; DB 1; Length 827;
Best Local Similarity 28.7%; Pred. No. 3.5e-38;
Matches 235; Conservative 127; Mismatches 318; Indels 140; Gaps 24;

113 ISPEHNSGHTTYDSCHIQYLLMGVLS-----SLTVPMQDQGLMGIMAHNSKPRRF 167
62 ILPTAEAD-RYRMD-----IGVLTGEEVSVEPI-----AVPGEVRL 101
168 TEQEMMALISK-----EVSIAITQSLSRQVHQOQVE-----ALVQLETT-- 211
102 VTR--KSRALITGSSSEKVIYALALDYTECRFAEALQASVEHNSILTEHPVPTAD 159
212 ----VAQGRPETWOYA-LETVGQAVEADGAVLYIAPDLTGSVAQHYQWNLRFDMQNL 266
160 PSGEVLEIGRPWEKTYAPKAEALGA-----GMKAMRPDLGVEYQ-----EAKSL 206

OY	267	ETSLMOELMRQPSRANERPMNAAYGS--TWKRPRTFSVAPRLPRNCVPHGYTL---	GELE 321
Db	207	AT-----GEPIDVEERFLAAAGGYSWRSRRATRRA-----	EDGSILIRWYGYE 250
OY	322	QRSDWIAPEPESLSENFQSFILVIVLADQOV-----	GSLILDR-----KEKSLVK 367
Db	251	DIDRRKMFELAKKSEARFRALADDAVYMLVINGENADADDYHSRLMETTGGTAEDQAGK 310	
OY	368	HWAGKRGIDRNNILPRLSFEAMETQOKLVPTWNSSEKLAQAVASTOLYMAITQQ-----	421
Db	311	GMLNAVHEDDDRNAVERFYQAFDLREPV-----RMEKRLRAGGSGSAMVIDIQPREASD 365	
OY	422	-----FV-----TRILITQQ-----TAYQPLQPLWMIIFNQTLALDAL 457	
Db	366	GTFLEGFVGLADLITERRAAEDEBRLLAQKQIHNAHRNHALTGRLR-ROFLKEEERERLSDH 424	
OY	458	YEGKMGVLVIAMDREKFRKINESFGHKTGDLLOEVDADRLNOKLSPLAAYSPLLSRHWG 517	
Db	425	APSTRLLATCLDLEGFKAINDAYGRAGRDLLIRVETRLRNFLEK---QSDILCRLSGDE 480	
OY	518	FTLILTOISDNQEMIPCEERLSTFQEPFFLOGQPIYLTASMSGTAPYQGETAESILKF 577	
Db	461	FVYLVRGVGINSNAERBLAQQILDIYEAPYELAGHVDQVYVGLAASKSDQSDLEIKT 540	
OY	578	AEIALTRAKCGCKMTYQGYEYRQDQSAFMLDRLTLESIDLQALTNQEFVLYRQPOVALDTC 637	
Db	541	ADIALERAKTGGGGITVQYEPRKMDADLRARQRMKVSRLHAKELSEVRQPLANLRTGQ 600	
OY	638	LLGVATLRMRQHPRLRGVAPRVFPLAEELGLINHLQOWLETPACAHQHFRETGRRLR 697	
Db	601	ITTFEALARKMHPERGOVSPAEPFLAVALDETEGLPDLSEWILROCTAYVM---PPYVS 656	
OY	698	MAVNISAFQFODEKWLNSVLECLARTQMPREDELLETETESIMMEDIKGTVLLRLHREEG 757	
Db	657	VAVNLSPLQFNRQRASTIVRSNLSLDTGDAISRQLLETITESVLLSECDOSNIQTLKEIRQLG 716	
OY	758	VQVAVIDEGTGYSSLSILKQPLPIRLKTIKDSFVNDLNEGADTAIIQYVIDLANGLNLET 817	
Db	717	VIVAVLDDEGTGYSSLSYLRTPPEFKIKVDRSFIDLKRSKESIALIVRAVAAGRISGLGIT 776	
OY	818	VAEGISEAQGLORLOKMGCHLGGCYFLLTRPLPAEAMMTYL 857	
Db	777	TVEGVERQDQDLTIKAEGFDAQGLTGGPLPASAQAMALL 816	
RESULT 4			
ID	YDDU_ECOLI	STANDARD:	PRT: 799 AA.
AC	P616129: P77708: P76872:		
DT	15-DEC-1998 (Rel. 37, Created)		
DT	15-DEC-1998 (Rel. 37, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Hypothetical protein yddu.		
GN	YDDU OR B1489.		
OS	Escherichia coli.		
OC	Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;		
OC	Escherichia.		
OX	NCBI_TaxID=562;		
OX	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-K12 / MG1655;		
RX	MEDLINE=97426617: PubMed=9278503:		
RA	Blattner F.R., Plunkett G. III, Bloch C.A., Berna N.T., Burland V.,		
RA	Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,		
RA	Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,		
RA	Mau B., Shao Y.;		
RT	"The complete genome sequence of Escherichia coli K-12.";		
RL	Science 277:1453-1474(1997).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-K12;		
RA	Alida H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,		

```

RA  Itoh T., Masai H., Kashimoto K., Kimura S., Kitakawa M.,
RA  Kitagawa M., Makino K., Miki T., Mizobuchi K., Mori H., Mori T.,
RA  Motomura K., Nakade S., Nakamura Y., Nishimoto H., Nishio Y.,
RA  Oshima T., Saito N., Sempel G., Seki Y., Sivasubram S.,
RA  Tagami H., Takeda J., Takemoto K., Takeuchi Y., Wada C.,
RA  Yamamoto Y., Horikuchi T.;
RT  "A 570-kb DNA sequence of the Escherichia coli K-12 genome
RT  corresponding to the 28.0-40.1 min region on the linkage map.";
RL  DNA Res. 3:363-377(1996).
CC  -1- SIMILARITY: BELONGS TO THE YEGE/YHDA/YHUK/YJCC FAMILY.
CC  -1- SIMILARITY: CONSTRAINTS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.
CC  -1- SIMILARITY: CONSTRAINTS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL, AE000246; AAC74562.1; ALT_INIT.
DR  EMBL, D90789; BAA15144.1; -.
DR  EMBL, D90790; BAA15154.1; -.
DR  EMBL, D90791; BAA15160.1; -.
DR  EcoGene: EG13792; yadU.
DR  InterPro: IPR001633; DUF2.
DR  InterPro: IPR000160; DUF9.
DR  InterPro: IPR001610; PAC.
DR  InterPro: IPR000700; PAS-assoc.C.
DR  InterPro: IPR000014; PAS.
DR  Pfam: PF00563; EAL; 1.
DR  Pfam: PF00990; GDEF; 1.
DR  Pfam: PF00989; PAS; 2.
DR  SMART: SM00267; DUF1.
DR  SMART: SM00052; DUF2.1.
DR  SMART: SM00086; PAC; 2.
DR  SMART: SM00091; PAS; 2.
DR  PROSITE: PS0112; PAS; 2.
DR  PROSITE: PS0113; PAC; 1.
KW  Hypothetical protein; Repeat; Complete proteome.
FT  DOMAIN 10 79 PAS 1.
FT  DOMAIN 134 205 PAS 2.
FT  DOMAIN 208 260 PAC.
SQ  SEQUENCE 799 AA; 90260 MW; 4CE770D2A6B39D5F CRC64;

Query Match 10.3%; Score 684.5; DB 1; Length 799;
Best Local Similarity 33.3%; Pred. NO. 1.le-36;
Matches 172; Conservative 99; Mismatches 204; Indels 41; Gaps 12;

QY 355 SLILRKESILYKHA-KGRGIDRNIIPLRSEFAMEETOK-----LVPTWNRSE 403
DB 292 SLFLNRNGPI--HMASSHGAELQNA-----QSMASATIRROGAPGIIQIKRSSGAE 343
QY 404 -----RKLAQVASTOLYMAITQOEYTRLTQOTFAVDPLTQLEPNMIIENRQTLALDLALY 458
DB 344 TSAETIERVADISQHAALALEQEKSRHIEQLQIEDPMWGRP-----KNNLNHLYLDLV 398
QY 459 ESKKMG--VLVIAMDREKRINSEFGHKIGDGLQEVADLNLKLSPLAAYSPLSRMGD 516
DB 399 D-KAVSPVYVILGVGHIDQVDSLGYNADQALLLEVNRFRKLKP-----DQYLCRIGCT 453
QY 517 GFTILLTQISDNQEMIPLEERLSTFOEPEFLQGPFIYLTASMGISTAPYD-GETAESLL 575
DB 454 QE-VLVSLENDVSNITQADELRNVSKRIMIDKPFLITISGIS---YDGKRRDYLL 509
QY 576 KRAEIALTRAKQCKNTQOFYRPQDSAPMLDLTLESDLRQALTNQOEFLYRQPVALDT 635
DB 510 STAHNAMDYIRKNGNGWQFSPANNENKRLVLGAALKRAISNNQKLKLYQPOIEAET 569
QY 636 GTLLVEALVVRQHGRLQGVADVFPLAEELGLINHLGQAWLLETRACATHQHFRETGR 695

```



```

DB 570 GELTGIEALARRHDP L HGVPPSRF L PLAEI GEIENIGMWIAEACROLAERNSONIHT 629
QY 696 LRMVAVNISAROFODEKMLNSVLECLRTGMPPEDELEETESLAMEDIKGTVLLHRLRE 755
DB 630 PALSVNLASLHFRSNOLPROMSDAMHANGIDGHOLVETREIMMEHDPEIFKRIQLRD 669
QY 756 BEVQVAVIDDEGCGYSSLSLTKOLPIHRLKIDKSFVNDLNEGADTAIOYVIDLANGLN 815
DB 690 MGVGLSVDFGFGFSGLSRLVSLPVEIKIDKSFVDRCLERKRIALLLEAITSIGOSLNL 749
QY 816 ETVAEGIESEAO L QRLQKMGCHLGOGYFLTRPPLAE 851
DB 750 TVVAEGVEKEDQEFEMKRIKHCRVIOGYFSPRLPAE 765

RESULT 5
YD54_MYCTU STANDARD: PRT: 623 AA.
AC 011024:
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 67.7 kDa protein RV1354c.
GN RV1354C OR MT1397 OR MTCY02B10.18C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriales; Mycobacteriaceae; Mycobacterium.
NCBI_TaxID=1773;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV:
RX MEDLINE-96295987; PubMed-9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badoock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN 12
RP SEQUENCE FROM N.A.
RC STRAIN-CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Kouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -! SIMILARITY: BELONGS TO THE YEGE/YHDA/YHKK/YJCC FAMILY.
CC -! SIMILARITY: SOME TO SYNECHOCYSTIS SP. SLR0359.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Z75555; CAA99987.1;
CC EMBL: AE007012; CAK45660.1;
CC HSSP: P03622; ZIFO.
CC TIGR: MT1397;
CC InterPro: IPR001633; DUF2.
CC InterPro: IPR000160; DUF9.
CC InterPro: IPR003018; GAF.

```

```

DR Pfam; PF00563; EAF; 1.
DR Pfam; PF01590; GAF; 1.
DR Pfam; PF00990; GGF; 1.
DR SMART; SM00267; DUF1; 1.
DR SMART; SM00052; DUF2; 1.
DR SMART; SM00065; GAF; 1.
DR Hypothetical protein; Complete proteome.
SQ SEQUENCE 623 AA; 67653 MW; ED54E5FA97191B7 CRC64;

Query Match 9.4%; Score 621; DB 1; Length 623;
Best Local Similarity 32.5%; Pred. No. 1e-32;
Matches 160; Conservative 85; Mismatches 212; Indels 36; Gaps 9;

QY 399 WNRSEKRLQAVAST---QLYMAITQOFVTRLTITQGRVADPDLQPMWIIINRQTLTALD 455
DB 152 WHEAEELNALTMTITLFRQVQARVAERKRLYLD---HDLTGL-----HNRALLQHD 203
QY 456 -ALYERK--MVGVLVIAAMDREKRINESFGHRTGDGLQEVADRNLKSLPLAAYSPLLSR 512
DB 204 QRLAPQPGPVALFLDIDLKAINDYLGHAAGDQFIHVAQRIGDAL----VGESLIAR 259
QY 513 WHDGDFTILLTQISDNQEMIPLEBRLLSTQPEFFLQGPFIYITASGISTAFYDGETAE 572
DB 260 LGSDEFFVLIPASPSADAAQPLAERLDQKDHVAIGGEVLTFTVTSIGVASGTPGQHTPS 319
QY 573 SILKFAEIALTRAKCOGKNTYQFRRPQDSAPMLDRLTLESDDLQALTNQEFVLYPOPOVA 632
DB 320 DLIRADQAAALAKHAGGSVALFTADMSVSGSLRNDIELHRLRGIESDLRLVLYPEVD 379
QY 633 LDTGKLGYEALVRQWHPRLGQVAPVFIPLAEGLINHLGQVLETACATHQHFRE- 691
DB 380 LRTGDIIVGTEALVRQWHPRLGGLAPGCFIPVASEINLAGELRWVLRAC---NEFSEW 435
QY 692 ----TGRRLMVAVNISAROFODEKMLNSVLECLRTGMPPEDELEETESLAMEDIKGY 747
DB 436 QSGAGLHDALLRINVASAGOLVYGGFVPAVDITGGHGLDASSVCLTEVNVVQDLHTAR 495
QY 748 VLLHRLREEGVOVAIDFGTGYSSLSLTKOLPIHRLKIDKSFVNDLNEGADTAIOYVY 807
DB 496 ATLRLKEGVHVAIDDFGYSALSILQTLPIDLTKIDKTFVRQGTNMSDLVIYRGIM 555
QY 808 DLANGINLETVAIGISEAO L QRLQKMGCHLGOGYFLTRPPLAEMMTYTYIQIILDFG 867
DB 556 TLAEGFQLDVVAEGVETEAARILLDQRCYRAGGFLESFVPEAMRHML-----S 606
QY 868 TPPLKVALPETE 880
DB 607 ARRLPPTCIPARD 619

RESULT 6
YCIIR_ECOLI STANDARD: PRT: 661 AA.
AC P77334:
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein yciir.
GN YCIIR OR B1285.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
NCBI_TaxID=562;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE-97426617; PubMed-9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";

```

```

RL Science 277:1453-1474(1997).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN-K12:
RX MEDLINE-97251357; PubMed-9097039;
RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itch T.,
RA Kasai H., Kasahimto K., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N.,
RA Samped G., Seki Y., Sivasubram S., Tagami H., Takeda J.,
RA Takeuchi K., Takeuchi Y., Mada C., Yamamoto Y., Horuchi T.,
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 28.0-40.1 min region on the linkage map.";
RL DNA Res. 3:363-377(1996).
RN [3]
RN SEQUENCE OF 1-557 FROM N.A.
RA Rita Z.R., Arraiano C.A.;
RL Submitted (FEB-1996) to the EMBL/Genbank/DDBJ databases.
CC -1- SIMILARITY: BELONGS TO THE YESE/YHDA/YHUK/YUCC FAMILY.
CC -1- SIMILARITY: CONTAINS 1 PAS (PER-ARNT-SIM) DIMERIZATION DOMAIN.
CC -1- CAUTION: REF.3 SEQUENCE DIFFERS EXTENSIVELY TO THAT SHOWN HERE DUE
CC TO MANY FRAMESHIFTS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000226; AAC74367.1; -
DR EMBL; D90767; BAA14847.1; -
DR EMBL; D90766; BAA14839.1; -
DR EMBL; L40788; -; NOT ANNOTATED_CDS.
DR EcGene; EG13415; yjcR.
DR InterPro; IPR001633; DUF2.
DR InterPro; IPR000160; DUF9.
DR InterPro; IPR000014; PAS.
DR Pfam; PF00563; EAL; 1.
DR Pfam; PF00990; GGDEF; 1.
DR Pfam; PF00989; PAS; 1.
DR SMART; SM00267; DUF1; 1.
DR SMART; SM00052; DUF2; 1.
DR SMART; SM00091; PAS; 1.
DR PROSITE; PS00112; PAS; 1.
KW Hypothetical protein: Complete proteome.
FT DOMAIN 109 179 PAS.
FT CONFLICT 4 4 V-> L (IN REF. 3).
FT CONFLICT 26 27 SS-> AA (IN REF. 3).
FT CONFLICT 113 113 A-> P (IN REF. 3).
FT CONFLICT 175 175 V-> L (IN REF. 3).
FT CONFLICT 256 256 D-> H (IN REF. 3).
FT CONFLICT 344 344 P-> G (IN REF. 3).
SQ SEQUENCE 661 AA; 74664 MW; 991D47CDE46A6F01 CRC64;

Query Match 9.2%; Score 607; DB 1; Length 661;
Best Local Similarity 31.6%; Pred. No. 9e-32;
Matches 144; Conservative 110; Mismatches 165; Indels 36; Gaps 8;

OY 401 RSEKLAOVASTQIWMATQOFVTRLLTQOTAYDPLTQLPNMIIFNRQLTALLDALY-- 458
DB 228 RAOERLRILANT-----DSITGLP-----NRNMODLIDHAINH 261
OY 459 -EGKAVGVIVMDREKRINESGKHTGDLLOEVDRLNOKLSPLAAVSPISRHHG 517
DB 262 ANNNKVVGVYLDLDNKKVNDAYGHLFGDQLRDVSLAI--LSCV-EHDQVYLRGSG 317
OY 518 FTLLTQISDNQEMIPICRLSTFQEPFLQOGPIYLTASMGISTAPYDGETAESLKF 577
DB 318 FLVLAISNTSQS-ALEMASRIILRLRPPRIGLIEVYTSQSVGIALSPRHGSDTAIRH 376

```

```

OY 578 AEIALTRACGCKNTYQFRRPDSAPMLDRLLLESDBRALNQNQEVLYFQPOVALDCK 637
DB 377 ADTAMTAKGGRGOCFCVTPENQVRFEYLMVDNLRLALENDQVLYHQPITW-RGE 435
OY 638 LIGVEALVWKPMPRLGQVAPDVFIPLAELGLINHLGQVAVLETACTHOFRETRRR 697
DB 436 VASLEALVWQSPERGLIPPLDPIISAESGLVPLGRVIT-LDVYQVAKRRDGINR 494
OY 698 MAVNISARQFODEKMWLSVLECLKRTGMPDELEETESLMEMEDIKGTVLLHRLREG 757
DB 495 VAVNISARQADQITFTALKQVQLNEFYCPIDVELTESCLIENDELALSVYQFSQLG 554
OY 758 VOVAIDDEFTGYSLSLTKQLPIHRKIDKSYNDLNGCATITAIQVYIDLANGINLET 817
DB 555 AQVHLDDEFTGYSSLSQLARFPIDAKLQVYVRDTHKQPVQSQSLVAIVAAQALNLQV 614
OY 818 VAEIGSEAOLORLQKMGCHLGQGYFLTRPLPAEA 852
DB 615 IAEVSEAKEDAFILKNGINERQGLFPAKMPRAVA 649

RESULT 7
YHUK_ECOLI
ID YHUK_ECOLI STANDARD; PRT; 651 AA.
AC P37649;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein yhjK.
GN YHUK OR B3529.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE-94316500; PubMed-8041620;
RA Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;
RT "Analysis of the Escherichia coli genome. V. DNA sequence of the
RT region from 76.0 to 81.5 minutes.";
RL Nucleic Acids Res. 22:2576-2586(1994).
RN [2]
RP IDENTIFICATION BY MASS SPECTROMETRY.
RX MEDLINE-94420866; PubMed-10493123;
RA Fountoulakis M., Takacs M.-F., Berndt P., Langen H., Takacs B.;
RT "Enrichment of low abundance proteins of Escherichia coli by
RT hydroxyapatite chromatography.";
RL Electrophoresis 20:2181-2195(1999).
CC -1- SIMILARITY: BELONGS TO THE YESE/YHDA/YHUK/YUCC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U00039; AAB18506.1; -
DR EMBL; AE000429; AAC76554.1; -
DR EcGene; EG12256; yhjK.
DR InterPro; IPR001633; DUF2.
DR InterPro; IPR000160; DUF9.
DR InterPro; IPR003660; HAMF.
DR Pfam; PF00563; EAL; 1.
DR Pfam; PF00990; GGDEF; 1.
DR Pfam; PF00672; HAMF; 1.
DR SMART; SM00267; DUF1; 1.
DR SMART; SM00052; DUF2; 1.
KW Complete proteome.
SQ SEQUENCE 651 AA; 73080 MW; 9C914999411D9E8D CRC64;

```

```

CC -1- SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X63841; CAA45329.1; ..
DR PIR; S18623; S18623.
DR HSSP; P23222; IDRM.
DR InterPro; IPR001633; DUF2.
DR InterPro; IPR000160; DUF9.
DR InterPro; IPR003018; GAF.
DR InterPro; IPR001610; PAC.
DR InterPro; IPR000700; PAS-aassoc_C.
DR InterPro; IPR000014; PAS.
DR Pfam; PF00563; EAL; 1.
DR Pfam; PF01590; GAF; 1.
DR Pfam; PF00990; GGDEF; 1.
DR Pfam; PF00989; PAS; 1.
DR SMART; SM00267; DUF1.
DR SMART; SM00032; DUF2.1.
DR SMART; SM00065; GAF; 1.
DR SMART; SM00086; PAC; 1.
DR SMART; SM00091; PAS; 1.
DR PROSITE; PS50112; PAS; 1.
DR PROSITE; PS50113; PAC; 1.
KW Hypothetical protein.
KW DOMAIN 185 PAS.
FT DOMAIN 265
FT SEQUENCE 735 AA; 80476 MW; CFF94DDF06440B9A CRC64;
SQ
Query Match 8.1%; Score 535; DB 1; Length 735;
Best Local Similarity 24.7%; Pred. No. 5, 2e-27;
Matches 221; Conservative 125; Mismatches 338; Indels 212; Gaps 28
QY 9 DFLRNINVKFRALITLRETQVYVEARIFLGVDRIYK-FPSDSGGEVLAENVRAAL 67
DB 32 DLLARIADTEH-----LQVSLFVHEHEGNG-IAALCVIDMR 69
QY 68 PSLIGLHPP-VEDIPQAREELGNORKMLAVDVARRKKSHEISGRIPTSHSNG----- 121
DB 70 P--GLAFPAPTESHPPLTAAGDP--LLAQMAARRRGETILGR---TRDITGLYIG 120
QY 122 ---HTTVDSCHIOYLLAMGVLSILYVPMODQOLGIMAVHNSKPRPRFEOEMETMAL 178
DB 121 FFHSH-----GVTFELTEPVVHNGRWG--HFCVDTPDHEHTVAVERQ 162
QY 179 SKEVSLATQSOQLRNQVHQOQVOEALVQRLFTTVAQYGDREPTQVYALFTVGAVEADGA 238
DB 163 AFKCIAAVLAGLILASGTEGLYSEPAARAM----- 192
QY 239 VLYIAPDLTGSVAQHYQWNLRFDMGNMLETSLMDQLMFGQPSAMEPAAVQSTWKP RP 298
DB 193 -----LDTSIDAYIVADEGALVEFNHAAEALFGHTRE 225
QY 299 FTSVAPLPPTNCVPBGYTLGLGEQSDMIAPRESISANFOSFLIPLPAADQOWGSLIL 358
DB 226 GVIGRPMTET-IIPAHYIDRHHQGFMRHATGENIMRRLYE-V EALRAD-----GS-V 276
QY 359 LKREKSLVHMAGKGRIDR---RNILPRLSEAMEETOKLYETNRSEERKLAQVASTOLY 415
DB 277 FPAELTVNHRHGRGLRFLFAFVRIDISDRIT-----SRRLERLAFBTH 320
QY 416 MAITQGF-VTRLTQGTAVDPLTQLEPNMILFNROLTIALLDLYGKMYGVAVIMDFK 474
DB 321 TGLSTNTGLRLTGLGRPT-----RSGAVVIMLRDQ 352
QY 475 RINESGCHTGGGLLEQVADRLNOKLSPLAAYSPLLSRWHDGFTILLQIISDNQEMIP L 534

```

```

DB 353 YVKTSGDDWAEPMIVETANLLSRMLPQEC-----LGRGSESEFTVYVQPGAAE---L 405
      : ||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 535 CERLLTFOEPFLOGOPILYTFASMGISTAPYDGETAESILKFAETALTRACOGKNYQ 594
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 406 AETLIGRLSALIESGRRYRLRGVVERPGD---ATYLLRDAEMA--ARDCROGHLH 460
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 595 FRRPDASAPMLDRLLTESDLRQALTNQ--EFVLYFOPQVALDTGKLLGVEALVRMHPRL 652
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 461 FAEHM-RACHOOOLEMLEMLRDVIORTTSALSHQPVVSARTGLVGEFALRWYSETH 519
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 653 GOVAPVFTPLAEELGLINHLGQWLET---ACATQHFEFRETG-PRRLMAVNISARQFQ 708
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 520 GPVSPALFPLAAGFAEELGAWVETALSACAGNNVRRANGLAPMHAIILNSTEVV 579
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 709 DEKMNSVLECLKRTGMPEDELEETESTLAMEDIKGTVYLLHRLREGVQVAILDFGFG 768
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 580 APDLIERVROTMAFHGLPQCVCFELETESAILNQPEIAIETLSRLALGCTTAIDFGFG 639
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 769 YSSLSTLKOLPIHRLKIDKSFVNDLNEGADTAIIQVVIDLANGLNLETAEGISEAOL 828
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 640 YSSLSTLQRLPMDVYLKIDNSFVLDVNDNSRREIVRMIMANGLGMSVABEVETTGAL 699
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 829 QRLQKMGCHLGQGYFLTRPLPAEAMTYLYYQIILDFGPTPLPKVALPETETEAG 884
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 700 QILRQMGCDRAQGFLEGRAMPQVAGT-----LPETLAPTG 735
      : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT- 9
YJCC_ECOLI STANDARD; PRT; 528 AA.
ID YJCC_ECOLI P32701; P76789;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein yjcc.
GN yjcc OR B4061.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=94089392; PubMed=8265357;
RA Blattner F.R., Burland V.D., Plunkett G. III, Sofia H.J.,
RA Daniels D.L.,
RT "Analysis of the Escherichia coli genome. IV. DNA sequence of the
RT region from 89.2 to 92.8 minutes."
RL Nucleic Acids Res. 21:5408-5417(1993).
CC -1- SIMILARITY: BELONGS TO THE YEGE/YHDA/YHJK/YJCC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U00006; AAC3155.1; -
DR EMBL; AE000479; AAC77031.1; -
DR Ecogene; EG11938; yjcc.
DR InterPro; IPR001633; DUF2.
DR Pfam; PF00563; EAT_1.
DR SMART; SM00052; DUF2; 1.
DR Hypothetical protein; Complete proteome.
KM SEQUENCE 528 AA; 60801 MW; F6E4819954912F31 CRC64;
SQ

```

```

QY 608 LTLFESDLRQALNDFVLYFOPQVALDTGKLLGVEALVRMHPRLGOYA-PDVEFIPLAE 666
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 269 LSPKKRLQALKEHOLCHYOPIDITKEKICGABALLRMGEO-GQIMNAEFLPLAEK 327
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 667 LGLINHLGQWLETACATHQHFRETGRRLR-----MAVNISARQFODEKWLNSVLECL 720
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 328 EGMIGQIDTVYD-----NVFRDLGDLATHADRYVINSASDPFTSRILARINOKT 380
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 721 KRTGMPEDELEETESTLAMEDIKGTVYLLHRLREGVQVAILDFGFGTSSSTLKOLPI 780
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 381 EQAVARPDQIKFEVTEHAFELVDKMTPIIL-AFRQAGTEVALDDGISTNLHNLKSLNV 439
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 781 HRLKIDKSFVNDLNEGADTAIIQVVIDLANGLNLETAEGISEAOLQRLQKMGCHLGQ 840
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 440 DLTKIDKSFVETLTTHKNSHLAEHIELASLGLKTAIEGVETEQVNLKRGVRYCQ 499
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 841 GFYLRPLPAEAMTYLYYQIILDFGPTPLPKVALPETETEAG 885
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 500 GWEFAKAMPQVFMQWM-----EQLPARELTRGQ 528
      : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT- 10
YLIE_ECOLI STANDARD; PRT; 782 AA.
ID YLIE_ECOLI P75800;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein ylie.
GN ylie OR B0833.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.,
RT "The complete genome sequence of Escherichia coli K-12."
RT Science 277:1453-1474(1997).
RL [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horinouchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map."
RL DNA Res. 3:137-155(1996).
CC -1- SIMILARITY: CONTAINS 1 DUF2 DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000185; AAC73920.1; -
DR EMBL; D90721; BAA35528.1; -
DR EMBL; D90722; BAA35536.1; -
DR Ecogene; EG13476; ylie.
DR InterPro; IPR001633; DUF2.

```

Query Match 6.0%; Score 400; DB 1; Length 528;
 Best Local Similarity 33.0%; Pred. No. 1.8e-18;
 Matches 94; Conservative 60; Mismatches 99; Indels 32; Gaps 6;


```

Db 587 PTVSQARFPEVSQLLAKYOIEAMQILFEVENSALTNTVKAQITLLOHDELGCQIAID 646
Qy 764 DEFGYSSLSLTKOLPIHRLKIDKSFVNDLNEGADTAIIQYVIDLANGLNETVAEGIE 823
Db 647 DFGYASARAKKNVADLTKIDGFSFINVSNLSDYOIVASICHARRKKMLVVAEYVE 706
Qy 824 SEAOQLRQKMGCHLGQGYFLTRPLP 849
Db 707 NEIREANVLISGIDYMGQYLTGKPP 732

RESULT 12
YD57_MYCTU STANDARD; PRT; 307 AA.
ID YD57_MYCTU 011027;
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 40, Last annotation update)
DE Hypothetical 33.9 kDa protein RV1357c.
GN RV1357C OR MT1400 OR MTCY02B10.21C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NGBL_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cone S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.,
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CC11551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson R.J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RA "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE YEGE/YHDA/YHCK/YJCC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: Z75555; CA99962.1; -
DR EMBL: AE007012; AAK45663.1; -
DR TIGR: MT1400; -
DR Tuberculist: RV1357c; -
DR InterPro: IPR001633; DUF2.
DR Pfam: PF00563; EAL; 1.
DR SMART: SM00052; DUF2; 1.
DR Hypothetical protein: Transmembrane; Complete proteome.
FT TRANSMEM 158 178 POTENTIAL.
FT TRANSSEM 203 223 POTENTIAL.
FT SEQUENCE 307 AA; 33982 MW; 3DC597A6AB3BAC48 CRC64;

```

```

Query Match 5.6%; Score 368.5; DB 1; Length 307;
Best Local Similarity 35.9%; Pred. No. 8.5e-17;
Matches 89; Conservative 46; Mismatches 106; Indels 7; Gaps 3;

Qy 610 LESDRQALTNOEFLVYPOVALDTGKILGYEALVRWQHPLGVAPVFIPLAEGL 669
Db 57 LSTSRVALDRGEFFLVYQPIRLADNRITGAEALRMEHPITGLTPRFIDRAENGL 116
Qy 670 INHLGQVLETCATGHOHFEFE--TGRRLRMVNISARQFOEKVNSVLECKTRTGMPP 727
Db 117 MYPLTAFVLEQAC-RHVSWMKRDHSTDPQFVSVNVSASTICDPGLVVEGVLGETGLPA 175
Qy 728 EDLELEITESLMEDIKGTVLLRLREGVQVAIDDEGTGYSSLTKOLPIHRLKIDK 787
Db 176 HALQLELAEDARLSDEKAVTRLQELSLAGVGAIDDEIGFSSLAIVPLRPDVVKLG 235
Qy 788 SFV----NDLNEGADTAIIQYVIDLANGLNETVAEGIESAOQLRQKMGCHLGQGYE 843
Db 236 KTECLDGDIOARLANEQITRAMIDLGKLTITVAKLVETPSQARLANAFCKAAQGMH 295
Qy 844 ITRPLPAE 851
Db 296 FAKALPVD 303

```

```

RESULT 13
YLAB_ECOLI STANDARD; PRT; 516 AA.
ID YLAB_ECOLI P77473;
DT 01-NOV-1997 (Rel. 35, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein ylab.
GN YLAB OR B0457.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RA "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ROBERTS D., Allen E., Araujo R., Aparicio A., Chung E., Davis K.,
RA Duncan M., Federapfel N., Hyman R., Kalman S., Komp C., Kurdi O.,
RA Lew H., Lin D., Namath A., Oefner P., Schramm S., Davis R.W.;
RA Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE YEGE/YHDA/YHCK/YJCC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AE000152; AAC73559.1; ALT_INIT.
DR EMBL: U82664; AAB40212.1; ALT_INIT.
DR Ecocore: EG14237; ylab.
DR InterPro: IPR001633; DUF2.
DR Pfam: PF00563; EAL; 1.
DR SMART: SM00052; DUF2; 1.
DR Hypothetical protein: Transmembrane; Complete proteome.
FT TRANSMEM 6 26 POTENTIAL.
FT TRANSSEM 242 262 POTENTIAL.

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 7, 2002, 18:53:41 ; Search time 60.43 Seconds

(without alignments)
3652.847 Million cell updates/sec

Title: US-09-272-809-2

Perfect score: 6614

Sequence: 1 NNPNRSLEDFLRNINKEFRR.....GRNQWLAYEGSQLPHWDEY 1276

Scoring table: BLOSUM62

Gapop 10.0 , Gapept 0.5

Searched: 562222 segs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL.19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriophage:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	922	13.9	1415	16	Q9HV18 pseudomonas
2	882	13.3	748	16	Q9A3V7 pseudomonas
3	865	13.1	1244	16	Q55576 synechocyst
4	862.5	13.0	1245	16	Q915W1 pseudomonas
5	858.5	13.0	772	16	Q92KJ3 rhizobium m
6	830.5	12.6	760	16	Q916K5 pseudomonas
7	811.5	12.3	783	16	Q9HYT3 pseudomonas
8	811	12.3	685	16	Q91310 pseudomonas
9	810.5	12.3	899	16	Q9HUF2 pseudomonas
10	809	12.2	687	16	Q9HW35 pseudomonas
11	801.5	12.1	840	16	P72843 synechocyst
12	797	12.1	1578	16	P74400 synechocyst
13	789.5	11.9	892	2	Q9L4F1 bacillus ce
14	785	11.9	724	16	Q984Y8 rhizobium 1
15	780.5	11.8	653	16	Q981Y4 rhizobium 1
16	774.5	11.7	808	16	Q98G38 rhizobium 1

17	769.5	11.6	614	16	Q98T06	Q98J96 rhizobium 1
18	768	11.6	951	16	Q9HTC4	Q9HTC4 pseudomonas
19	758.5	11.5	774	2	Q9AF11	Q9AF11 vibrio para
20	757	11.4	685	16	Q9K065	Q9K065 vibrio chol
21	752.5	11.4	856	16	Q9RS28	Q9RS28 deionococcus
22	746	11.3	864	16	Q91243	Q91243 pseudomonas
23	739.5	11.2	732	16	P73380	P73380 synechocyst
24	734.5	11.1	635	2	Q935Y8	Q935Y8 synechococ
25	731.5	11.1	634	2	Q93J80	Q93J80 streptomyce
26	731.5	11.1	842	16	Q9A666	Q9A666 caulobacter
27	717.5	10.8	564	16	Q92WP4	Q92WP4 rhizobium m
28	714	10.8	742	16	Q930X2	Q930X2 rhizobium m
29	714	10.8	1051	16	Q92YNO	Q92YNO rhizobium m
30	705.5	10.7	805	16	Q9RX49	Q9RX49 deionococcus
31	703.5	10.6	749	16	Q943J1	Q943J1 bacillus su
32	700.5	10.6	809	16	Q9ABX9	Q9ABX9 caulobacter
33	699.5	10.6	746	2	Q92584	Q92584 streptomyce
34	689	10.4	752	2	Q87376	Q87376 acetobacter
35	687	10.4	543	16	Q9K8N4	Q9K8N4 bacillus ha
36	686	10.4	853	16	P72746	P72746 synechocyst
37	678	10.3	735	16	Q9A9V4	Q9A9V4 caulobacter
38	677.5	10.2	772	16	Q97L01	Q97L01 clostridium
39	672	10.2	558	16	Q9HT09	Q9HT09 pseudomonas
40	671.5	10.2	446	16	Q983J9	Q983J9 rhizobium 1
41	670	10.1	714	2	Q9RCY2	Q9RCY2 streptomyce
42	665.5	10.1	738	16	Q98E82	Q98E82 rhizobium 1
43	663.5	10.0	554	16	Q985V5	Q985V5 rhizobium 1
44	663.5	10.0	814	16	Q67220	Q67220 aquifex aeo
45	661	10.0	740	2	Q87378	Q87378 acetobacter

ALIGNMENTS

RESULT	ID	Q9HV18	PRELIMINARY:	PRT: 1415 AA.
AC	Q9HV18	01-MAR-2001 (TREMBLrel. 16, Created)		
DT	01-MAR-2001	(TREMBLrel. 16, Last sequence update)		
DT	01-DEC-2001	(TREMBLrel. 19, Last annotation update)		
DE	HYPOTHETICAL PROTEIN PA4601.			
GN	PA4601.			
OS	Pseudomonas aeruginosa.			
OC	Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;			
CC	Pseudomonas.			
OX	NCBI_TaxID=287;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-ATCC 15692 / PA01;			
RX	MEDLINE-20437337; PubMed-10984043;			
RA	Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,			
RA	Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,			
RA	Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,			
RA	Brody L.L., Coulter S.N., Folger K.R., Kas A., Lartig K., Lim R.M.,			
RA	Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,			
RA	Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;			
RT	"The complete genome sequence of Pseudomonas aeruginosa PA01, an			
RT	opportunistic pathogen.";			
RL	Nature 406:959-964(2000).			
DR	EMBL; AE004874; AAC07989.1; -			
DR	InterPro; IPR001633; DUF2.			
DR	InterPro; IPR00160; DUF9.			
DR	InterPro; IPR003015; HLR_Myc.			
DR	InterPro; IPR001610; PAC.			
DR	InterPro; IPR000014; PAS.			
DR	InterPro; IPR000700; PAS-assoc_C.			
DR	Pfam; PF00990; GCDERF.1.			
DR	Pfam; PF00785; PAC.3.			
DR	SMART; SM00267; DUF1.1.			
DR	SMART; SM00052; DUF2.1.			
DR	SMART; SM00086; PAC.3.			
DR	SMART; SM00091; PAS.4.			

	Query Match	13.1%	Score 865;	DB 16;	Length 1244;
	Best Local Similarity	31.0%	Pred. No. 3	3e-51;	
	Matches 259;	Conservative 141;	Mismatches 286;	Indels 150;	Gaps 28
QY	145	VPVMDQ-----QLWGMAY-----	NHSPRRF-----	TEDEMTAL-----	177
		: : : :	: : : :	: : : :	
Db	475	IPFVADQTEITLTPLDSSMVAEMRVSPIHMGKAFPLSRDITTEGQARLALAESEK			534
		: : : : : : : : : : : : : : : :			
QY	178	-----LSKEVSLATISQSLSRVHQQ-----	QVQALQRLRETTYAQQGDRETFMQ		223
		: : : : : : : : : : : : : : : :			
Db	535	KRYIVELTSEGIWLLDQDQQTFFANQDLADMLGYSVQVETLERKNITAFVLVIHHEPSON			594

QY	224	YALFTV-----	GAQVEA----	DAVLYIADPLTGSVAQONLFFDGMN--	265
Db	595	SHQKTLDSFPCVLPVPHNGQYVDVQFORRDSVLM-----	GLVYSAMTD--	QMGNTRG	645
QY	266	---LETSLMOELMGOPSAMER----	MAAVOS-TWEKPREFTSAPRLPTNCYPHGVT	316	
Db	646	ELAVLDTITRKSAEQALASSEQRLBGLIGSIDDVWASDAVSPATLYLNPTTAAVYQGS		705	
QY	317	LGELEQSD-W---IAPESLSAENFOSFLIYPLAADQWGSLLIRKESKLVKNHAGK	372		
Db	706	LEVYQSONWFEEVYHGFDDLLETHYLDL--	MEKDO-----	TELEYRIVQSGKE	754
QY	373	RGIDRRNILDR-----	LSPEAMEEPOKLVPTWNSREKILQAVASTQYLMATTOGFVTRL	426	
Db	755	RWLRRSQVLVDGEGHOPRLRIDSIDS-----	DIETRKLA-----	AEK	790
QY	427	ITQOATADPDTLPQNPWTFIFNRQTLALDLALYEKGK-VGVLYIAMDREFKRNESFGHKTG	485		
Db	791	LHYAANHDSLNLNPNRSMFLDRIGMLQRLRRDRDLRFVALLFDLDEFFIINDLSIGHSGC	850		
QY	486	DGLLEQVADRLNOKLSPLAAYSPILSRHWGDEFTILTITQISNOEMIRLCEBLLSTFOER	545		
Db	851	DLLDQGIANHRLROCLRP----	EDTLARLGGDEFTMLIENITCPEDEVIAVACRIHOELOKP	906	
QY	546	FFLQGGQIYULASMGIS-TAPYDGEFAESLTKRAEALRLRACQCGKNTYQFRRPDSAPM	604		
Db	907	FNLNGQIEFTNTSISGLALNPNHGH--	PODVLRADTAMRAAKAAGGRAIFNQTMHNA	965	
QY	605	LDRLTLESDDRQALTNQNEFYLQFOPVADDTGKLYEVLAVMQRHQLQGVAPDVPIRLA	664		
Db	966	VQRLQRENDRLRAIDRQELDHLQPIYCLKTQGLQGVELAVLQKMRNEBSLLPREFVAIA	1025		
QY	665	EEELGLNHLQGVULETA----	CATHQFFRETEGRRLMAVNISAROFODEKVLNSVLSCL	720	
Db	1026	EETGLIYPMGDWILMEASRQILELKQSPOLS--	HLQVSINVSRLQRLQKLTVDLTL	1083	
QY	721	KRTGMPREDLELETESLIMEDIKGVVLLHNRREGCVQVAILDDDTGTSISLILKQRLP	780		
Db	1084	SSTMLAOPDKLETETTESLTLIDNLNLAADVLSLRKNQIDISLDDDTGTSISLSTLYHREP	1143		
QY	781	HLRLKDSFPNDLINESADPAIIQYVIDLANGENTVEAGIESENAQORLOKMGCHNQ	840		
Db	1144	NTIKVDSFVPTNMEPNQNNQAIYHTIYTLAHTLGLDVABEIGETERHILTOQLHMLGCDAGQ	1203		
QY	841	GVELTLPAPAMKTYLYYQIOLDFSPPTPLPKVALPETETBAGGQNVGDRPLPNS	896		
Db	1204	GYPFARPIPEDLLDFL-----	TLQIPRMSVPAV-TECH-----	GDRKEPFS	1244
RESULT	4				
Q915W1		PRELIMINARY;	PRT;	1245 AA.	
AC	0915W1				
DT	01-MAR-2001	(Tremblrel. 16, Created)			
DT	01-MAR-2001	(Tremblrel. 16, Last sequence update)			
DT	01-DEC-2001	(Tremblrel. 19, Last annotation update)			
DE	HYPOTHEICAL PROTEIN PA0575.				
GN	PA0575.				
OS	Pseudomonas aeruginosa.				
OC	Bacteriia, Proteobacteria; gamma subdvision; Pseudomonadaceae;				
OC	Pseudomonas.				
OX	NCBI_Taxid=287;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
KC	STRAIN-ATCC 15692 / PA01.				
RX	MEDLINE=20437337; PubMed=10984043;				
RA	Stover C.K., Plam X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,				
RA	Hickey M.J., Plam X.-O.T., Hufnagle W.O., Komalik D.J., Lagrou M.,				
RA	Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,				
RA	Brody L.L., Coulter S.N., Polger K.R., Kas A., Ladbai K., Lim R.M.,				
RA	Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,				
RA	Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.,				
RT	*Complete genome sequence of Pseudomonas aeruginosa PA01. an				

QY 577 FAELALTAACGCGKATYFYRPODSAPMLDLRLTLESLSLRQALTNQEFVLYTPOVALDTG 636
D 487 NADMAAYAKELKRNIVTEEMAAARAEKIQSOEELREALAREPFLHPPOKMLETG 546
QY 637 KLIGVEALVRMOHPRLTGOVADVFIPLAEELGLINHLGOWLETACAHQHFFETGRRL 696
D 547 RIFPAEALLRNQHPRGRIISPAFIPLAEFGLIYPIGDWLTATC-----ROL 595
QY 697 R-----MAVNISAROFODEKWLNSVLECLRTGMPEDLEETESLAMEDIKG 745
D 596 KAMODAGPLIVSVNARQFERBNMAARVAALVEETCLAREYLELETSLINQDVP 655
QY 746 TVVLIHLRREGVOYADDEGTGSSLSLKOLPIHLKIDKSFVNDLNEGADPAIIOY 805
D 656 AIAIMHLEALGVLAIDDFGTGYSLSALKRFPVRLKIDRSFVTIDIPHDADMAITSA 715
QY 806 VIDLANGLELVEAGIESEAOLRLQKMGCHLGOGYFLTRPLPAE 851
D 716 IISLAQKLGRLVIAEGVETEAQVEFLQKSGCEIGYFFSOPLPE 761

RESULT 6
QY16K5 PRELIMINARY: PRT: 760 AA.
AC QY16K5;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HYPOTHETICAL PROTEIN PA0285.
GN PA0285.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hutnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltzy L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Ladbis K., Lm R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sater M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
DR EMBL: AE004466; AAC03674.1; -
DR InterPro: IPR001633; DUF2.
DR InterPro: IPR000160; DUF9.
DR InterPro: IPR001610; PAC.
DR InterPro: IPR000014; PAS.
DR InterPro: IPR00700; PAS-assoc_C.
DR Pfam: PF00990; GDEF; 1.
DR Pfam: PF00785; PAC; 2.
DR Pfam: PF00989; PAS; 1.
DR SMART: SM00267; DUF1; 1.
DR SMART: SM0052; DUF2; 1.
DR SMART: SM00086; PAC; 2.
DR SMART: SM00091; PAS; 2.
KW Hypothetical protein; Complete proteome
SQ SEQUENCE 760 AA: 86380 MW: 049AFDIDEE222E05 CRC64:

Query Match 12.6%; Score 830.5; DB 16; Length 760;
Best Local Similarity 33.3%; Pred. No. 3.8e-49;
Matches 238; Conservative 104; Mismatches 246; Indels 125; Gaps 18;

QY 214 QYGRPEFTWYAL--ETVGQAVEADGAVLYIAPDLTGSVAOHYQWNLRPDCN--WLETS 269
D 120 ELGSDFTWKRLHPDEEYATVLEAHRNH--QGTLNDNDIHYR--LRHKDDDYWINS- 173
QY 270 LMQELMRQPSAAMPMAVYSTWKEPRPTSVAPLRPTNCVPHVGTJGELG----- 322

D 174 -----RGR-----VLRLALCKPLHYTGA-----RITTLQRLKEDHROAA 209
QY 323 -----RSDMI---APPESIAENFOSF---LIV 344
D 210 AVFSDTREGVLTDAQAVIVHNVPSEFRTITGYSRSDVGLKFTALLRSQDQDAFQRLML 269
QY 345 PLADQQWVGSLLILRKESLVKHNAGKRGIDRRNIPRLSEAMEETOKLYPTNNRSE 404
D 270 ALREDDVMSGGEIWNRRKSGEIVPQWLHRAV--RNDQGLTGYVGVFSD--LSISIKRSN 325
QY 405 KLAQVASTQLYMAITQQFTRITTOATADPLTQLPNWIIFNRQELTALLDLYEGKMG 464
D 326 ELDFLAH-----HDSLTGLPNRYLREIREQALENKDKRTVACA 364
QY 465 VLVIMDRFKRINSEFGHKTGDLLQEVADRNLQKSLPLAAVSPILSRHNGGFTILLTQ 524
D 365 LLILDLHFKHINDSLGHTGMDLKEYSKRQHQDENC-----LLSRIGSGDFALV-E 419
QY 525 ISDNQEMIPLCERLLSTQEPFLQGPITYLTASMGISTAPYDGETAESLKPFAITAT 584
D 420 NDDPEAVARLSQRIIDGFNAPPDICHQPIYISASIGSVLPEDASDVHDLMQADALFQ 479
QY 585 AKCGCKNTYQFYRPODSAPMLDLRLTLESLSLRQALTNQEFVLYTPOVALDTGKLGVEAL 644
D 480 AKDSGRMAAFYTRVLTARAHVQVESALRHLEHDELRYHPVHDLASGRIVGESL 539
QY 645 VMQHPRLGOVADVFIPLAEELGLINHLGOWLETACAHQHFFETGRRLR-MAVNI 703
D 540 VMQHPRLGOVADVFIPLAEELGLINHLGOWLETACAHQHFFETGRRLR-MAVNI 598
QY 704 ANQFQDEKWLNSVLECLRTGMPEDLEETESLAMEDIKGTVLLHRLREGVOYAD 763
D 599 SRLFNNGGLEERIANALRESGLEPRYLETEVESAMEDEFSQSLNLRITLGVLAID 658
QY 764 DGTGYSLSLKOLPIHLKIDKSFVNDLNEGADPAIIOYVIDLANGLELVEAGIE 823
D 659 DGTGYSLSLKOLPIHLKIDKSFVNDLNEGADPAIIOYVIDLANGLELVEAGIE 718
QY 824 SEAOQLRLQKMGCHLGOGYFLTRPLPAEAMTYLYLPQIDRGPTPLKVALPE 878
D 719 HDDQALFLREHCGDGGQGYWGRPQPALEALR-----DLRPAALPE 759

RESULT 7
QYHT3 PRELIMINARY: PRT: 783 AA.
AC QYHT3;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HYPOTHETICAL PROTEIN PA3311.
GN PA3311.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hutnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltzy L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Ladbis K., Lm R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sater M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
DR EMBL: AE004466; AAC03674.1; -
DR InterPro: IPR001633; DUF2.
DR InterPro: IPR000160; DUF9.

DR InterPro: IPR000215; Serpin.
 DR Pfam: PF00990; GDEF; 1.
 DR SMART: SM00267; DUF1.
 DR SMART: SM00052; DUF2; 1.
 DR PROSITE: PS00284; SERPIN; UNKNOWN.1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 783 AA; 86815 MW; 7D81B379356430F CRC64;

Query Match 12.3%; Score 811.5; DB 16; Length 783;
 Best Local Similarity 29.7%; Pred. No. 8.5e-48;
 Matches 239; Conservative 119; Mismatches 285; Indels 161; Gaps 19;

QY 160 HSKRRR-----TEQETMALLSKE-----VSLATQSLRQVHOQ 198
 DB 43 HRRRRVRFVQGDFTAMDQGLRFLGESPYDGYLQNCYSPSLVALAFVLAAGTALD 102
 QY 199 QVQVALVORLETTVAQYGRPEWTQ-----VALETV 229
 DB 103 -----MVERGNSLS-----HPLMOWIGAFCGSGIMATHEVAMLAFAHPIALRDLPT 153
 QY 230 GOAVEADGAVLYIAPDLT-----GSVAOHYOMNLFDMGNWLETSL 270
 DB 154 GLSLIIAASVLTWMTARPRFGLPCLLACCGIGIAAHY----- 197
 QY 271 WOELMRGQSAAMEPMAAVQSTWEK-----RPFVSAPLPPTKCVPHGYTLG 318
 DB 198 -----TGMAMRSVATQYQPSLFSVLIAIGAAFTALAVP-----YLRG 239
 QY 319 ELEORSQWIAPEPSL-----SAENFQSP-----LIVPLAADQWQGSLLLRKESLYKH 368
 DB 240 RRSARYRWKLIASLLGALIAAHFTGMALVLSVPACT-----PLELQASADSLRIG 293
 QY 369 WAGKRGIDRRNLPRLSEAMEETQKLVPTWNRSEKLAQVASTOLYMAITO-QEVTRLI 427
 DB 294 WL--TGVLASAIAACGIAAAMEKOR-----ERRLSE--NSSVNALNLQDHAHSL 341
 QY 428 TQQTAYDPLTQPMIIFNRQTLTALLDALYRGKMGVLYVIMDKFRKRNESFGKRTGDG 487
 DB 342 ROMARYDSLTLGQNTAATNEVFVQHLNCRKGLKGLAVFLLDHFKRINDSLGHDSGQ 401
 QY 488 LLOEVAIDLNOKLSPLAAYSPLLSRHMGDFITLLTQISDNQEMPLCERLSTFOEPFE 547
 DB 402 LKTIYSERIRSVLRD-----SDVVARFADGECVLAIDLQDHEAHI--LSQRLMKKKKEPIA 456
 QY 548 LQGDPIYLTASMGISTAPYDGETAELSLKFAEIALFRACCKQNTYOFYRPQDSAPMLDR 607
 DB 457 LDGRTLVMTASGVSLYPNDGECCELLKNGALHQSACGRNNAOFFSRQLLVRAIOE 516
 QY 608 LTLBSDLQALTNQEFVLYFQPOVALDPTGKLLGVBALVRMQRRLGOVAPVFTLAEEL 667
 DB 517 LQMBELKQALRDQDLHYQPTIALADGEVHQLEALVRMRPRTGGLGPDREIGLAEAN 576
 QY 668 GLINILGQWLETACATQHFFRETRGRRLMAVNISARQFODEKMLNSVLETKRTGMP 727
 DB 577 GMIDQDLQWVLRACRDLRSLHAGHERLRYAVNCCASNGLRASLVDEVRHLEBAGLAA 636
 QY 728 EDLELEITESIMEDIKTQVYVLLHRLREGVOVALIDEGTGYSSLSLKQDPIHRLKIDK 787
 DB 637 CFLELEVETEDALNYIDQITPLERLRELGVSLSIDFGTGYSSLAYLRRLDLALKYDR 696
 QY 788 SFVNDLNEGADTALIQVIDLANGLNEYAEGIESERQOLQKMGCHGCGYFLTRP 847
 DB 697 SFINDIPASQDMEIAQAIIMAAQRLHLKVVAAEGVETPOOLAFLRENHCEIVGYSFRP 756
 QY 848 LPAEAMTYLYYPOIDGCPTRPL 871
 DB 757 LPLALEEFL--RAYRPDAAPL 777

RESULT 8
 Q91310
 ID Q91310 PRELIMINARY: PRT: 685 AA.

AC Q91310;
 DT 01-MAR-2001 (TREMBLrel. 16; Created)
 DT 01-MAR-2001 (TREMBLrel. 16; Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19; Last annotation update)
 DE HYPOTHETICAL PROTEIN PA1727.
 CN PA1727.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 OX NCBI_TaxID=287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 15692 / PA01;
 RX MEDLINE=20437337; PubMed=10984043;
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
 RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
 RA Gaeher R.L., Goltz L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Lathin K., Lim R.M.,
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 RA Reizer J., Salter M.H., Hancock R.E.W., Lory S., Olson M.V.,
 RA "Complete genome sequence of Pseudomonas aeruginosa PA01, an
 RT opportunistic pathogen.";
 RI Nature 406:959-964(2000).
 DR EMBL; AE004599; AAC05116.1;
 DR InterPro: IPR001633; DUF2.
 DR InterPro: IPR000160; DUF9.
 DR Pfam: PF00990; GDEF; 1.
 DR SMART: SM00267; DUF1.
 DR SMART: SM00052; DUF2; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 685 AA; 74378 MW; 4BF74F5CE494A0C7 CRC64;

Query Match 12.3%; Score 811; DB 16; Length 685;
 Best Local Similarity 34.4%; Pred. No. 7.4e-48;
 Matches 196; Conservative 107; Mismatches 119; Indels 88; Gaps 13;

QY 340 SFLIVPLAADQ-WV-----GS--LILLRREKSLV-----KHWACK----- 372
 DB 145 SLILVIVASGALMTAFRLRHGSRRLVLRGALVMGCAIVGMHYTMAAOPFGSFC 204
 QY 373 ----RGIDRRNLPRLSEAMEETQKLVPTW-----NRSEKLAQVASTOLY 415
 DB 205 GAAGGCID-----NGMLAVLYVITTLAVIAIALYVLSRLER--TSVATLSLA 253
 QY 416 MAITQOYFVRLTQGTADPITQPLPMIIFNRQTLTALLDALYEGKMGVLIAMDREKR 475
 DB 254 RA-----NRELQALHDNLTKLPNRMLDRLDRLQQAIRQDRRAVLFMDLGEKA 307
 QY 476 INESFGHTGSGDLQEVADRLNOKLSPLAAYSPLLSRHMGDFITLLTQISDNQEMPLC 535
 DB 308 VNDVAGHNLGDLLEVAERIRANVRA---QDTIARLGDEF-VLLIAREPDAATYA 362
 QY 536 ERLSTFOEPPEFLOQPYLTLASMGISTAPYDGETAELSLKFAEIALFRACCKQNTYOF 595
 DB 363 EKLVRKRIQPYQIISHNEVRISISGIALYPRDQGRHMLNADAMNAADQGRNGYCF 422
 QY 596 YRPQDSAPMLRRLTLESDLQALTNQEFVLYFQPOVALDPTGKLLGVBALVRMQRRLGOV 655
 DB 423 FESSNNANAAQQLDQLHDLRLQALERQVLYHYQPRVLAAPNPMGYEALLMEHPHGLI 482
 QY 656 APDVFIPLAEELGLINILGQWLETACATQHFFRETRGRRLM-----AVNISA 704
 DB 483 TPQGFPLAEKTYGLIVQGEWVDEAC-----RQRLMIDGGHADMIANVLSA 531
 QY 705 RQFODEKMLNSVLECKTKTGMPPREDELEITESIMEDIKTQVYVLLHRLREGVOVALD 764
 DB 532 LQFAHAGLVDSVRALLRHSLEPSHLLIEVRESAMRADASVILRLQSLAMGVGISTD 591
 QY 765 FGTGYSSLSLKQDPIHRLKIDKSFVNDLNEGADTALIQVIDLANGLNEYAEGIES 824
 DB 592 FGTGYSSLSLYKRLPASLTKIDRGFINELAHSDDAALVSAIVALGRTLNKLYABGEVET 651

```

QY 825 EAQJOLKMGCHLGQGYFLTRPLPAEAMM 854
DB 652 EAQJELTRIGCNLSLOGFLGRPMPEQLL 681

RESULT 9
Q9HUF2 PRELIMINARY: PRT: 899 AA.
AC Q9HUF2;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HYPOTHEITICAL PROTEIN PA5017.
GN PA5017.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
CC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
  Garber R.L., Goltzy L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
  Brody L.L., Coulter S.N., Folger K.R., Kas A., Latdig K., Lam R.M.,
  Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
  Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
  "Complete genome sequence of Pseudomonas aeruginosa PA01, an
  opportunistic pathogen.";
  Nature 406:959-964(2000).
  RT Nature 406:959-964(2000).
  RL EMBL, AE004914; AAG08402.1; -.
  DR InterPro: IPR001633; DUF9.
  DR InterPro: IPR000160; DUF9.
  DR InterPro: IPR003018; GAF.
  DR InterPro: IPR001610; PAC.
  DR InterPro: IPR000014; PAS.
  DR InterPro: IPR000700; PAS-assoc.C.
  DR Pfam: PF01590; GAF; 1.
  DR Pfam: PF00990; GDEF; 1.
  DR Pfam: PF00785; PAC; 1.
  DR SMART: SM00267; DUF1; 1.
  DR SMART: SM00052; DUF2; 1.
  DR SMART: SM00065; GAF; 1.
  DR SMART: SM00086; PAC; 1.
  DR SMART: SM00091; PAS; 1.
  DR Hypothetical protein; Complete proteome.
  KW Hypothetical protein; Complete proteome.
  SQ SEQUENCE 899 AA; 101321 MW; 925CIFA936731602 CRC64;

Query Match 12.3%; Score 810.5; DB 16; Length 899;
Best Local Similarity 27.5%; Pred. No. 1.3e-47;
Matches 248; Conservative 151; Mismatches 289; Indels 215; Gaps 27;

QY 5 RLEDFELRNVLNK-----FHALT---LRETLQVYIEEARIFLGVNRVYIKRA 50
DB 160 RSQDHLQHLKSRTOQNLVRLANRHYLSDDLLEAQLTQACAEAVGTARAGIWRLL 219
QY 51 SDGSGEVL-----EAVNRALPFLGLHFPVEDIPPQAREELGNQRKMIADV 98
DB 220 DDQRLNAVTVRRDLQYEEKPSIDASRYPAL-----EAVNSGALIDNH 264
QY 99 VAHRKKKHEHLSGRISPTSHSNGHTTVDSCHIQVILANGVLSLTVPMQDQGLWIMA 158
DB 265 NAOBRPFRQEL-----YKDYLRPLGVNALLDATIRIGEVVGYLC 304
QY 159 VVHSPRRFTQEWETMLLSKEVS-----LATTSQSLSQVNOQVOQVLELVRLTTYAAO 214
DB 305 LEHAENFM-----WO-----SDEIAFAGELADQYAOVLNHNHRRVSSLL----- 345
QY 215 YGDRRETFQVAVLTVGQA---VEADGAVLYIAPDLTGSVAQYQNNLRPDMGNMLETSL- 270
DB 346 -----HLQRAVVEQSASAFLLIDRDGVVEYVNPST-SITQVSADEVR-----NRRLSELPL 395

```

```

QY 271 ----WQELRGQPSAAMEPMAAVOSTWEKRPRTSVAPLPPTNCVPHGYTLEGEQRSDM 326
DB 396 ALENISELLFDARSALTQ-----QNSMQ-----GEFRSR----- 424
QY 327 IAPPELSAENQSFILVPLAADQOVGSIILIRKESL--VKHNAKKGIDRRNRLPRL 384
DB 425 -----RKNHEPY-----W-GQLSLSKYVDJGELHYIG----- 452
QY 385 SPEAMEEQKLVPTWRSEKRLAQVASQTLYMAITQGFYRLTTOOTAY-DPLTOLPMNI 443
DB 453 -----IYEDITQNKLAQOHTKLAIRDNITGLA---- 480
QY 444 IENKQTLALDLALYBG---KNVGVLYIAMDRKRNESFGHKTGDLQEVADRNLQRL 500
DB 481 --NRHYFIALERLESSQDRPLSLILVINDNFKRINDSLQGTQDKLVSLARRLSCL 538
QY 501 SPLAATSPILSRWHDGFTILL--TQISDNQEMIPICERLSTFQEPFLQOPILYLTAS 558
DB 539 GGGAT---LARPASNEFVLLDDTAVERGES---IAQVLIHMLDKPLFVDNQLINITGS 591
QY 559 MGISTAPYDGETAESILKFAETALTRAKCGKNVYGFYRPOSAPMLDRLTLESRLQAL 618
DB 592 IGLASAPQGCPOPTLMKTAGLAKAKANGKQOVYFEALTAEKSTYLPESNLRRL 651
QY 619 TNOEFLYRPOVALDTRGLGVEALVRWQHPRLGQAVDFEPLAEELGLNHLGQWVL 678
DB 652 AQNELLAHVHQPKLCILRSQGLLEALLRMQHPEKMIKRPDRFISVAEFTGLYIPDKWVI 711
QY 679 ETACATHQHFFRETGRRLMAVNIASROFODEKUNSVYELCKRTGMPDELETTESL 738
DB 712 REACRQARELAERAGLGELOIAINLSRPTDPLVGSIAALHEENIPASOLELETTESL 771
QY 739 NMEDIKGYVLLHRLREEGVOVAIDPFGTGSLSLTKQPLFHLRLKIDSFVNDLNEGA 798
DB 772 LLDATDDTRQOELRLKSLGTLTAMDFTGYSLSLTKFPIDYIKIDSFIKIDIPSDQD 831
QY 799 DTAITQVYIDLANGLLETFVAEGISEAQOLRLQKMGCHLGQGYFLTRPLPAEAMMT-YL 857
DB 832 DMEITSAVIYAMHNKLKVAEGVESAEQOLAFIRNRCDIGOGYFLDRPISDLNTLSLL 891
QY 858 YYP 860
DB 892 RYP 894

RESULT 10
Q9HW35 PRELIMINARY: PRT: 687 AA.
AC Q9HW35;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)
DE HYPOTHEITICAL PROTEIN PA4367.
GN PA4367.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
CC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
  Garber R.L., Goltzy L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
  Brody L.L., Coulter S.N., Folger K.R., Kas A., Latdig K., Lam R.M.,
  Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
  Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
  "Complete genome sequence of Pseudomonas aeruginosa PA01, an
  opportunistic pathogen.";
  Nature 406:959-964(2000).
  RT Nature 406:959-964(2000).
  RL EMBL, AE004852; AAG07755.1; -.

```

DR InterPro: IPR001633; DUF2.
 DR InterPro: IPR00160; DUF9.
 DR Pfam: PF00563; EAL; 1.
 DR Pfam: PF00990; GDEF; 1.
 DR SMART: SM00267; DUF1; 1.
 DR SMART: SM00052; DUF2; 1.
 DR Hypothetical protein; Complete proteome.
 KW HYPOTHETICAL PROTEIN; COMPLETE PROTEOME.
 SQ SEQUENCE 687 AA; 77364 MW; 4907154941BC53DE CRC64;

Query Match 12.2%; Score 809; DB 16; Length 687;
 Best Local Similarity 35.8%; Pred. No. 1e-47;
 Matches 226; Conservative 85; Mismatches 248; Indels 72; Gaps 16;

QY 266 LETSMOELMR-----GQPSAMEPMAAVOSTWEKRPPTSVAPLPPTNCV----- 311
 DB 75 IEGLFQHEAVROASIGHPG---EPMLA-----EKSRLPDL-----PTRWLTDPILGQERT 122
 QY 312 -----PHGYTIGELQQRSDMTAPRESLAENF--QSFLVPLAAQOQWGSILLR 360
 DB 123 PSIRLGRPPYSEYGDLTITLD-TAP---YGENVTTSITLIFISGLRALAMGLVLF- 176
 QY 361 KEKSLVKNHAGRGIDRRNLRPL-SFEAMETOKLVPT---WNSEKRLAQVASTOLYM 416
 DB 177 ----LVYHNMILKPLSK--IIEHLVSINDRPSQHOLPLKHERNELGLMTTANQLLA 230
 QY 417 AITQOQVTRL-----ITQOATVPLTOLPWNIIENKQLTALLDALYES-----KMGVL 466
 DB 231 STESNHLRREAEDMLRTISQYDFLTGLP-----NQQLLOQDLDQILDGAGQORRVAVL 285
 QY 467 VIANMRFRINSEFGHKQDGLQEVADRLNOKLSPLAAYSPLLSRMHNDGFTILLTQIS 526
 DB 286 CIGLDFEKGINQYVQLGDQDLIALADRLRGHSARLGS---LARLGDGDFALVQADIE 341
 QY 527 DNOEMIPCELRITSFQEFPLQOGPIYLTASMGISTAPYDGETAESLFAEIALTRAK 586
 DB 342 QPEYAEALQSLDLDLEAFETIDQNEHVRKATIGITLPEDEBTEKLLQKAEQMTLAK 401
 QY 587 CQCKTQYFYRPQDSAPMLDRLLTESDLRQALTNQOEFVLYFQROVALDITGKLLGEALVR 646
 DB 402 TRSRNRQYFYIYASVSEMRRELEKDLRQALRHEHLNVQYQYDVRHVVGEALLR 461
 QY 647 WQHRPLGOVARDVFIPLAEELGLNHLGQWVLETACATHQNHFFRETGRLLRAVNI SARQ 706
 DB 462 WQHRPLGQVARDVFIPLAEELGLNHLGQWVLETACATHQNHFFRETGRLLRAVNI SARQ 706
 QY 707 FODEKMLSVLECLRTGMPREDLELETESLAMEDIKGTVVLLRLREEGVQVADIDFG 766
 DB 522 LHHNMLPRVNSMLQVYLPARASLELEVTETGIMEDISTAQNHLSTLRAGALLAIDDFG 581
 QY 767 TGYSSLSILKQLPRLHRLKIDSFNVDLLNEGADTAIIQYVIDLANGLELTVAESIESEA 826
 DB 582 TGYSSLSYLSKSLPLDKIKRIDSFNVDLLNEGADTAIIQYVIDLANGLELTVAESIESEA 826
 QY 827 QLOLRQKMGCHLGGYFLRLPRLPAAAMTYL 857
 DB 642 QEAYITAEQCNQGYLYSKPLPARELTQYL 672

RESULT 11
 ID P72843 PRELIMINARY; PRT; 840 AA.
 AC P72843;
 DT 01-FEB-1997 (TREMblrel. 02, Created)
 DT 01-FEB-1997 (TREMblrel. 02, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE NITROGEN FIXATION POSITIVE ACTIVATOR PROTEIN.
 GN NIFL OR SLR1305.
 OS Synechocystis sp. (strain PCC 6803).
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
 RN NCBI_TaxID=1148;
 RX [1]
 RP SEQUENCE FROM N.A.

RA MEDLINE-97061201; PubMed-8905231;
 RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
 RA Miyajima N., Hiroseawa M., Sugitara M., Sasamoto S., Kimura T.,
 RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naru K., Okumura S.,
 RA Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
 RA Tabata S.,
 RT "Sequence analysis of the genome of the unicellular cyanobacterium
 RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
 RT entire genome and assignment of potential protein-coding regions.";
 RT DNA Res. 3:109-136(1996).
 CC -I- SIMILARITY: THE N-TERMINAL REGION IS SIMILAR TO THAT OF OTHER
 CC REGULATORY COMPONENTS OF SENSOR TRANSDUCTION SYSTEMS.
 CC EMBL: D90901; BAI16858.1; -.
 DR HSSP: P08402; 1B00.
 DR InterPro: IPR001633; DUF2.
 DR InterPro: IPR00160; DUF9.
 DR InterPro: IPR001610; PAC.
 DR InterPro: IPR000014; PAS.
 DR InterPro: IPR000700; PAS-assoc.C.
 DR InterPro: IPR001789; Response-reg.
 DR Pfam: PF00563; EAL; 1.
 DR Pfam: PF00990; GDEF; 1.
 DR Pfam: PF00785; PAC; 1.
 DR Pfam: PF00989; PAC; 2.
 DR Pfam: PF00072; response-reg; 1.
 DR SMART: SM00267; DUF1; 1.
 DR SMART: SM00052; DUF2; 1.
 DR SMART: SM00086; PAC; 2.
 DR SMART: SM00091; PAS; 2.
 DR SMART: SM00448; REC; 1.
 KW Complete proteome; Phosphorylation; Sensor transduction.
 SQ SEQUENCE 840 AA; 96226 MW; 5F56F3D367BD985 CRC64;

Query Match 12.1%; Score 801.5; DB 16; Length 840;
 Best Local Similarity 28.2%; Pred. No. 4.8e-47;
 Matches 234; Conservative 150; Mismatches 268; Indels 179; Gaps 24;

QY 140 LSSLT--VPVMDQQLMGIMAVHNSKPRFTEQEMETNALSKEYSLAITQSLSQRYQH 197
 DB 86 LSALTQDSYVGLGLAVDYIH--KP--FQGE-----ITRQLQDKLKH- 127
 QY 198 QOVQALYQRLTETTYAQGDREPTWQALLEYGAQVADGAVLY-IAD----- 245
 DB 128 -QTNQALLQKNOLEBOIEKTAQTQALMOS-----EINFAVAFNOSPDPFIYIGRNSGC 181
 QY 246 -----LTG-SVAQHYQWNLRFDMGNUL-ETSLMOE----- 273
 DB 182 IMDVQRPQCPFGGLTKQELVGVSRQGFYFWVDEQOROSFLRDLMLQKDGLYKREKRE 241
 QY 274 -----LMRGQ-----SAMESPMA----- 288
 DB 242 VYDQEKQVMTLSGEPFENRVDCLLFVMRDIFERRAKQKLTLSQACQSPASIVIT 301
 QY 289 -VQS--TWKRRPPTSVAPLPPTNCVPHGYTLGELQFSMDIAPRESAENFQSFILVP 345
 DB 302 DVQGNITVYVNP-K-FEIS-----GYKSAEVLGNPRLISGNTQDYE-LMKKT 349
 QY 346 LAADQWVGSLLILKREKSLVKNHAGRGIDRRNLRPLSFEAMETOKLVPTWNRSRK 405
 DB 350 LASGRNHGFEHNRKKNEL-YW-----ERASIPISN----- 381
 QY 406 LAOVASTOLYMAITQOFV-----TRLITQOATVPLTOLPWNIIENKQLTALLDALYES 460
 DB 382 --QGIIVHYAAVKEDIKKEQQAELFNQAHYHGLTFLPRILAKDLQAISALNOK 439
 QY 461 KMGVLYIAMDREKRINESFGHKTGQGLQEVADRLNOKLSPLAAYSPLLSRMHNDGFTI 520
 DB 440 HIFGLMFLDLNFKKVNDFLTGHGADQDLVEVSETRAL-----RQDTVVARLGDEFLLI 495
 QY 521 LITQISDQEMIPCELRITSFQEFPLQOGPIYLTASMGISTAPYDGETAESLFAEIALTRAK 580
 DB 496 ILDOVSHSRKLMAITQRLRLRVMPVNLQGLEFFVHSGIGITVEPDDFFHADVILRNADT 555


```

QY 581 ALTKACQCKNTYQYFVROSAPRLDLETTESDRLALYNQEVLYAPROVALDTEKLG 640
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 556 AMYAAKLAGRMNEFFETPRHNAQAQMAIESELROSLKQEBOLYQAPVLSLESGOYG 615
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 641 VEALVRMOPRLGQVADVFETPLAEELGLINHLGQVWLFLACATNHHFFRETGR- LRMA 659
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 616 AEALMRHNRLLGTVPPDQGFPIPIEEGILLVLEGEMTDLVNCCQAAMHSHALGEQFWMS 675
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 700 VNISARODECKMNLNVELECKRKQMPDEPLETETESLMEEDIKGVVLLNHLREGVQ 759
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 676 VNVSERQKDSYFAIIQGFQKQVPRWEMLEETENLLLEENGDLKRLSLEENEA 735
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 760 VAIDFQGYSSLSLKQPLRIHLKIDKSFVNDELINAGDAPLAIQVYIDIANGLNETYA 819
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 736 LSLDDFGYSSLYLNRKPNFNSLKDIDRSFVELLPHDNTNVTGLVRAIIMAHHLKLYIA 795
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 820 EGISESQQLQRLQKMGCHLQGYLYLTPRLAEAMMYLYLYPQLIDRGPTTP 870
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 796 EGITLPRQMNFLRLOGCDYGGGYFESPAIAVEWES-----LGQRP 837
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT	12
P74400	
ID	P74400
AC	P74400; PRELIMINARY; PRT; 1578 AA.
DT	01-FEB-1997 (TREMBLrel. 02, Created)
DT	01-FEB-1997 (TREMBLrel. 02, last sequence update)
DT	01-OCT-2001 (TREMBLrel. 18, last annotation update)
DE	HYPOTHEETICAL 178.5 KDA PROTEIN.
GN	SL0267.
OS	Synechocystis sp. (strain PCC 6803).
OC	Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX	NCBI_TaxID=1148;
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=97061201; PubMed=8905231;
RA	Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA	Miyajima N., Hirosewa M., Suglira M., Sasamoto S., Kimura T.,
RA	Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA	Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M.,
RA	Taketa S.;
RT	"Sequence analysis of the genome of the unicellular cyanobacterium
RT	Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT	entire genome and assignment of potential protein-coding regions.,"
RL	Drug Res. 3:109-136(1996).
DR	EMBL; D90914; BAA18497.1; -
DR	InterPro: IPR001633; DUF2.
DR	InterPro: IPR000160; DUF9.
DR	InterPro: IPR003018; GAF.
DR	InterPro: IPR001610; PAC.
DR	InterPro: IPR000014; PAS.
DR	InterPro: IPR000700; PAS-assoc_C.
DR	Pfam: PF00563; EAL; 1.
DR	Pfam: PF01590; GAF; 1.
DR	Pfam: PF009980; GDFEF; 1.
DR	Pfam: PF00785; PAC; 2.
DR	Pfam: PF00989; PAS; 2.
DR	SMART; SM00267; DUF1; 1.
DR	SMART; SM00052; DUF2; 1.
DR	SMART; SM00065; GAF; 1.
DR	SMART; SM00086; PAC; 2.
DR	SMART; SM00091; PAS; 2.
DR	Hypothetical protein; Complete proteome.
QO	SEQUENCE 1578 AA; 178546 MW; BE1604B60FBA56DF CRC64;

Query Match	12.1%	Score 797;	DB 16;	Length 1578;
Best Local Similarity	29.8%	Pred. No. 2.8e+46;		
Matches 222; Conservative	124;	Mismatches	158;	Gaps 23;

QY 138 GVLSLTPVMOQOOLMGIAVHHSKPRFTEQEMETMALISKEVSLAITQSLSRQVHQ 197
| | | | : : : : | : | : : : | : | :
| | | | : : : : | : | : : : | : | :

Db	952	GFLLSSAAVPIRROGQVYAVLTFYSQAGFSEDEDLQLEISDNLSFALMADLQDDA--N	1009
QY	198	QOQVEALVQRLETTVAQYGDREPTEWQYALETVGQAVEADGAVLYIAPDLTSSVAAOHQWN	257
Db	1010	QRAON-----OLGENEERLRALAEANOG-----FYDLN	1038
QY	258	LRFPMGWMLTSLMQLMRQPSAAAEPMMAAOSTWEKRPFTSVAPLRPNCPHGHTL	317
Db	1039	LQ-----TGO--AVVSPQ-----YGGIL	1054
QY	318	G-----ELEQRSDWIAPEESLSAENFOSFLIVPLAADQWGSLLIRKESLIVKMA	370
Db	1055	CYDPQYQFQHEQR--WL---ERHPQ-----DQ-----KREIYDQYIQ	1089
QY	371	GKRGIDRNLLPRLSFEAWETQ-----KLVPWNSEKRLAQVASTOLYMAIT	419
Db	1090	GK-----IPRAVECRQRTWAGQWMLSLUGKIT--EMDSQGRPLMLG---ILVDVT	1137
QY	420	QOQFTRLITQTA--YDPLTQLPWITIFRNQTL-----ALDALYEGKMGVLYIADRF	473
Db	1138	ERKQAEQIENLNYDPTALP-----NRLLLDRAENMLAQRSKHFGAIVLIDIGF	1192
QY	474	KRISEFGHTGGGLLOEVADRILNOKLSPLAAYSPLSSWHDGFTILLISDNQEM--	531
Db	1193	KTLDANGHDSDBRLQMAAKRLADSLRD---SDTVAHGDEFITLLPELHAHRELA	1248
QY	532	---IPLERLSTFOEPPFLOGQPIYLTAASMGISTAPYUGFAEBSILKFAEIALTRAKQ	588
Db	1249	RLGIVGEKIRQALAMPFLTEAAQYQISSGITLTFPKLNENVSDFKFAQDTAMYGAKKA	1308
QY	589	GKNTYQFYRPQDSAPMLDRLTLESRLQALJNOEFVLYFOPOVALDTSKLLGVEALYRWQ	648
Db	1309	GRDVCLEFESQOMLEVESRFALFADRLSALOEKOPQVYLQPOVD--SHGVMIGACALLRN	1367
QY	649	HPRKGVAPDVFPLAEELGLIHNHGOWLVEFACATHQHFR--ETGRRLMAMVNISARQ	706
Db	1368	HPRGGLFPIPIFPIAEELIGIGIDFVLEOVC---QYLAFLQDLGLTRIALINVSPO	1424
QY	707	FODEKMWNSVLECLKRTGMPDEDELEIETESLMEDIKGTIVLLHRLREESQVVAIDFG	766
Db	1425	FQRAQNFAPKEIKNLACGVDPYRLTLEVEGLIVEDTHRALATMPELQTLGIHRSVDDFG	1484
QY	767	TGYSSLSILKQPLIHLKIDKSFVNDLNEGADTAIQQYVIDLANGLLEFVABGISSEA	826
Db	1485	TGYSSLSYLRRLPLNELKIDRAVYQAPQDLNNMALVEALISVARTFMALIAVEGVEETE	1544
QY	827	QLORLQMGCHLQGGYFLRPLPAE	851
Db	1545	QVOFLAQDGNFYQGYFGYGMPIID	1569

RESULT	ID	PRELIMINARY;	PRT;	892 AA.
09LAF1	09LAF1			
AC	09LAF1;			
DT	01-OCT-2000 (TREMBlrel. 15, Created)			
DT	01-OCT-2000 (TREMBlrel. 15, Last sequence update)			
DT	01-DEC-2001 (TREMBlrel. 19, Last annotation update)			
DE	YKOW PROTEIN.			
GN	YKOW.			
OS	Bacillus cereus.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;			
OC	Bacillus/Staphylococcus group; Bacillus.			
OX	NCBI_TaxID=1396;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=ATCC 14579 TYPE STRAIN;			
RX	MEDLINE=20055637; PubMed=10589720;			
RA	Ostad O., Gominet M., Purnelle B., Rose M., Kleclius D., Kolsto A.B.,			
RT	"sequence analysis of three Bacillus cereus loci under P1CR-regulated			
RT	genes encoding degradative enzymes and enterotoxin."			
RL	Microbiology 145:3129-3138(1999).			
EMBL	AF243712; CAB69812.1; -			

DR InterPro: IPR001633; DUF2.
 DR InterPro: IPR001610; DUF9.
 DR InterPro: IPR001610; PAC.
 DR InterPro: IPR000014; PAS.
 DR InterPro: IPR000700; PAS-associ_C.
 DR Pfam: PF00990; GDEF. 1.
 DR Pfam: PF00785; PAC. 1.
 DR SMART: SM00267; DUF1. 1.
 DR SMART: SM00052; DUF2. 1.
 DR SMART: SM00086; PAC. 1.
 DR SMART: SM00091; PAS. 1.
 SQ SEQUENCE 892 AA; 102911 MW; 1D49A6B1D150022E CRC64;

Query Match 11.9%; Score 789.5; DB 2; Length 892;
 Best Local Similarity 37.3%; Pred. No. 3.6e-46;
 Matches 174; Conservative 97; Mismatches 183; Indels 13; Gaps 4;

QY 395 LVPTW-NSSEKLAQVAS--TOLYMAITQOEVYTRILTOQTAYDPLTOLPNIIFNRQLTL 451
 Db 431 LIPPEINKEVGMFGIADITTLTKOKO-----VEHLAFHDALTLGPNRRKFEKDLKA 484
 QY 452 ALLDLYEGKMGVGLVIAMDRKINESGHTGDIQEVADRLNQLSPLAAYSPLLS 511
 Db 485 ILNTRQTNANDYAVIFLIDIRKKNIDRGLGELGDLLEIAKRLGCLRS---KDIYA 540
 QY 542 RMHGDGFTILLTQISDNOEMIPLCERLLSTQEPFLOGQPIYLTASMGISTAPYDGETA 571
 Db 541 RQGGDEFILLPEMYSEKSAIFIAQIITLILNKPFQIGELSTIPISIGIAMYPDGTDV 600
 QY 572 ESILKEFATILTRACOCKNTQYFRPODSAPMLDRILLESRLQALNQEVLVFPQYV 631
 Db 601 NELKNAAMAMRAKANRKNRFFSKESIAQNEIQLEGLSLALQONEFFLEYQPV 660
 QY 632 ALDTGKLLGVEALYKQPRIGQVAPADVFIPLAEELGINHIGVWLETACATHOFFERE 691
 Db 661 STKTQIIIGFEALIKWKPKKGIYSPAQFIPLAEETGFIIEGNILKTRACLEARWNNQ 720
 QY 692 TGRIRLVANVISARQFQDEKMLNSVLECLKRTGMPPEDELEITSLMEDIKGTIVLIH 751
 Db 721 GFSLHKGVLNLSVQFNHADIPLTISKVLETELPEALDEITETSIAINQSVAKLE 780
 QY 752 RLREGVQVADIDFGTSSLSILKQPLRIHRLKIDKSFVNDLNGADPAIIQVYIDLAN 811
 Db 781 QLQNLGIDISIDDFGTGSLAYLTQYPIINTKIAREFIGTGTSPLEAIVSITITLTK 840
 QY 812 GLNLETVAGIESAQLOLRQKMGCHLGQGYFLTRPLPAEAMTYLY 858
 Db 841 QLNLEVIAGVETEEMFEFLKQNCNDHIQGFISKPVSSKQVWMLLH 887

RESULT 14
 Q984Y8 PRELIMINARY; PRT; 724 AA.
 AC Q984Y8;
 DT 01-OCT-2001 (Tremblrel. 18, Created)
 DT 01-OCT-2001 (Tremblrel. 18, Last sequence update)
 DT 01-OCT-2001 (Tremblrel. 18, Last annotation update)
 DE ML7787 PROTEIN.
 GN Rhizobium loti (Mesorhizobium loti).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group.
 OC Phyllobacteriaceae; Mesorhizobium.
 RN NCBI_TaxID=381;
 RX STRAIN=MAFF303099;
 RX MEDLINE=21082930; PubMed=11214968;
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
 RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
 RA Takeuchi C., Yamada M., Tabata S.;

RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 Mesorhizobium loti."
 RL DNA Res. 7:331-338(2000).
 DR EMBL: AP003012; BAB54175.1;
 DR InterPro: IPR001633; DUF2.
 DR InterPro: IPR000160; DUF9.
 DR InterPro: IPR000014; PAS.
 DR Pfam: PF00563; EAL. 1.
 DR Pfam: PF00990; GDEF. 1.
 DR SMART: SM00267; DUF1. 1.
 DR SMART: SM00052; DUF2. 1.
 DR SMART: SM00091; PAS. 2.
 KW Complete proteome.
 SQ SEQUENCE 724 AA; 80637 MW; 0DF43B43CE5F39CB CRC64;

Query Match 11.9%; Score 785; DB 16; Length 724;
 Best Local Similarity 37.6%; Pred. No. 5.3e-46;
 Matches 183; Conservative 93; Mismatches 175; Indels 36; Gaps 7;

QY 368 HWAGRGIDRRNII--PLTSFEAMEEQKLVPTNRSERKLAQVASTQLYMAITQOEVY 424
 Db 253 HSAVRYLDGRTVIVTRQPMSSGGWVTFEDVTERRRTERMIHLAH----- 299
 QY 425 RLITQOTAYDPLTOLPNIIFNRQLTLALLDLYEGKMGVGLVIAMDRKINESGHT 484
 Db 300 -----HDTLGLPNSMFRERLIDLALDA--AAPLAIFSLDIRKAINDWGHPA 349
 QY 485 GDLGLQEVADRLNQLSPLAAYSPLSRWNGGFTILLTQISDNOEMIPLCERLLSTQ 544
 Db 350 GMLIKSAVERLQRCIH--NETDYVARGGGEFFVYVQNSGIADAELAKRIYEALAK 406
 QY 545 PFLQGPRIYLTASMGISTAPYDGETAESLKFATILTRACOCKNTQYFRPODSAPM 604
 Db 407 QPRDSRDMHIGISIGIAVFPODGRDADTLKNADTALYRAKSEGRNLYRFEPPADAIV 466
 QY 605 LDRLLLESRLQALNQEVLVFPQYVADLTGKLLGVEALYKQPRIGQVAPADVFIPLA 664
 Db 467 QARRALEVDLALPRQEDLDQPIIMTASGEIIGAELEKMWHSFARGTAPDAFIPVA 526
 QY 665 EELGINHIGVWLETACATHOFFETGRIRLVANVISARQFQDEKMLNSVLECLKRTG 724
 Db 527 EETGLIVPLGEMALKKACAAASW----PGLRIAVNNSAVQKGGFARSYISALASG 582
 QY 725 MPPEDELEITSLMEDIKGTIVLIHRLREGVQVADIDFGTSSLSILKQPLRIHRLK 784
 Db 583 VPAGOLELEITFTVLMDSKAVLKTIRLDIGIRIALDDFGTSSLSIAYLRFPVDKIK 642
 QY 785 IDKSFVNDLNGADP-ALIQVYIDLANGLNETVAGIESAQLOLRQKMGCHLGQGYF 843
 Db 643 IDRSFIRDMGNR--DTAAIVRTITGLGNELGIVTAEGVETEVDMLDRNDCBEAOGYL 700
 QY 844 LTRPLPA 850
 Db 701 IGVPSKA 707

RESULT 15
 Q981Y4 PRELIMINARY; PRT; 653 AA.
 AC Q981Y4;
 DT 01-OCT-2001 (Tremblrel. 18, Created)
 DT 01-OCT-2001 (Tremblrel. 18, Last sequence update)
 DT 01-OCT-2001 (Tremblrel. 18, Last annotation update)
 DE MLR9186 PROTEIN.
 GN Rhizobium loti (Mesorhizobium loti).
 OC plasmid pWLA.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group.
 OC Phyllobacteriaceae; Mesorhizobium.
 RN NCBI_TaxID=381;
 RX STRAIN=MAFF303099;
 RX MEDLINE=21082930; PubMed=11214968;
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
 RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
 RA Takeuchi C., Yamada M., Tabata S.;

RC STRAIN-MAFE0303099;
 RX MEDLINE-21082930; PubMed-11214968;
 RA Kaneo T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
 RA Watanabe A., Ideesawa K., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,
 RA Takeuchi C., Yamada M., Tabata S.,
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 RT Mesorhizobium loti.";
 RL DNA Res. 7:331-338(2000).
 DR EMBL: AP003015; BAB54575.1;
 DR InterPro: IPR001633; DUF2.
 DR InterPro: IPR00160; DUF9.
 DR Pfam: PF00563; EAL; 1.
 DR Pfam: PF00990; GGDEF; 1.
 DR SMART: SM00267; DUF1; 1.
 DR SMART: SM00052; DUF2; 1.
 KW Plasmid: Complete proteome.
 SQ SEQUENCE 653 AA; 71042 MW; 250B1D28BC5E61B8 CRC64;

Query Match 11.8%; Score 780.5; DB 16; Length 653;
 Best Local Similarity 32.0%; Pred. No. 9.2e-46;
 Matches 199; Conservative 109; Mismatches 228; Indels 85; Gaps 13;

QY 262 WGNMLET--LMOELMRGQPS-AAEPMAAYOSTWEKPRPFTSVAPLPTNCYPHGYTIG 318
 DB 65 WGSPTQTKALWASYVIGAGFVGLRARSFQSV--KPR--SVSRRTQNLVRNAFLFG 119
 QY 319 ELEORSQDWIAPP-----ESLSAENFOSFLIPLVPLADQO---WVGS 355
 DB 120 -----TWGALVLPFGAATNAQVYITCLSAQMTNGAASFTTIPAAIAVITLPIFVGS 174
 QY 356 LILREKSLVHWAGKRGIDRRNILLPLSFAMEETOKLPTWNRSEKLAQVASTOLY 415
 DB 175 AVAI-----VMEGAVNLPV-----ALIVSYAITLF 201
 QY 416 MAI-----TQOFTRLTTOOT-AYDPLTOLPMIIFNRQTLALLDALYEGKMGVLY 467
 DB 202 RAVLAHASEFTOSFILQTESENAIRRDVLTSLPNRPSFNERLENALVDARQEDHALL 261
 QY 468 IAMDREKRLNESGHTGGLQEVADRNLNOKLSPLAASPLSRWNGDFTILLTOISD 527
 DB 262 PDLNSNDEVDHFGRAMADAVLEMAARLRK-----TRESGIALRLEGGEFATIAARDIR 317
 QY 528 NOEMIPLCERLLSTFOEPFLQGOPIYLTAASGISTAPYDGETAESLKFPAETALTRAC 587
 DB 318 PQIGSLAKQIIDVMRAPFLIEGREIYCRYSVIGIALAPTDGLDANQLRCVPTALHRAKT 377
 QY 588 QGKNTYQFRRPDSAPMLDRLTLESPLRQALTNQEFVLYFOPOVALDITGKLLGEVALRW 647
 DB 378 LAGSTIQFSSASDDDAARRNALERDLASALANDEMLAFQEPFLDGSERIRGEFALLRW 437
 QY 648 OHPRLGQVAPDVEIFPLAEELGLINHGQVAVLETCACATHOHFFRETGRRLRMAVNISAROF 707
 DB 438 QHPTLGAIIPSEFIPIAETALIHITGHAAVKTACLAAYHWPRD---LRVSNLSAVOL 493
 QY 708 ODEKMLNSVLECIKRTGMPDEDELEETESLMEIDIKGTVLLHRLREGVQVAYIDDFGT 767
 DB 494 KDKALLDGIWALSETGLPEKRLVEVETESLISDFEALISLLQSLSPSLSVVALDDFGT 553
 QY 768 GYSSLSILKQPLIRHKLIRKISFVNDLINCAGADTAIIQYVIDLANGINLETVAEGISEAQ 827
 DB 554 GFSSTLYLTKPLSRKIDRSFVQMDLADACAIYRSIVOLAHLELRIVETAEGVETPEQ 613
 QY 828 LORLOKMGCHLGQGYFLTRPL 848
 DB 614 LDYLRVGCDEAQTIGKRPV 634

THIS PAGE BLANK (USPTO)


```

FT      /note= "histidine kinase transmitter module
FT      conserved motif"
XX      W09805944-A1.
XX      12-FEB-1998.
XX      01-AUG-1997; 97WO-US13529.
XX      02-AUG-1996; 96US-0023217.
XX      (REGC ) UNITV CALIFORNIA.
XX      Lagarias JC, Murphy JT;
XX      WPI, 1998-145711/13.
XX
XX      Adducts of apoprotein polypeptide and chromophore as label,
XX      particularly for bio-molecules - used as fluorescent markers in
XX      immunoassays, nucleic acid hybridisation, detecting protein-protein
XX      interaction etc., are stable with high molar absorption
XX
XX      Example 4; Fig 10B; 87pp; English.
XX
XX      This polypeptide comprises cyanobacterial phytochrome 1 (Cph1) of
XX      Synechocystis sp. PCC6803. Its amino acid sequence was deduced
XX      from locus sl10473 genomic DNA. Expression of the 748-residue
XX      polypeptide in E. coli and incubation with phycocyanobilin yields
XX      an adduct with a red, far-red photo-reversible phytochrome
XX      signature. The invention provides a new class of fluorescent
XX      protein adducts (designated phytofluors) that are generally suitable
XX      for use as fluorescent markers. They comprise a protein component
XX      (an apoprotein) and a bilin chromophore such as phycoerythrobilin.
XX      Preferred apoproteins are obtained from plants, e.g. oats (see
XX      AA50144), from green algae, e.g. Mesotetium caldariorum (see
XX      AA50145), or cyanobacteria such as Synechocystis. Truncated
XX      apoproteins consisting of the N-terminal chromophore domain are
XX      especially preferred. Recombinant apoproteins assemble
XX      spontaneously with the bilin chromophore. Claimed compositions
XX      comprise a protein, glycoprotein, antibody or nucleic acid to be
XX      detected linked to the fluorescent adduct. They are used in assays
XX      for detecting the other member of a specific binding pair, e.g.
XX      immunoassay of antigens, immuno-histochemical labelling, as nucleic
XX      acid probes for Southern blotting, for identification of manufactured
XX      products, also to detect protein-protein interactions, including
XX      studies on intracellular protein localisation and identification of
XX      transfected cells. The phytofluors make ideal fluorescent markers
XX      because they have a long wavelength absorption maximum and high
XX      molar absorption coefficient, and are stable to light and pH.
XX
XX      Sequence 748 AA:
XX
XX      Query Match 28.5%; Score 292; DB 19; Length 748;
XX      Best Local Similarity 36.8%; Pred. No. 8.5e-26;
XX      Matches 64; Conservative 38; Mismatches 64; Indels 8; Gaps
XX
XX      2 LAVRAISRLQSLPGDGICALCTVVEDVQRLTGYDRWVYGFHEDDHGEVSEIRSDLE 61
XX      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX      Db 139 manaaInIrIrg--gaInIdrfdydIvvevrrmIqfgyvImIrdennhcvIdakdrdme 196
XX
XX      62 PYLGHHYPRMTPQPARFLPFKONRVMICDN--APPVAVOSEELKRLCLVNSTLRP 119
XX      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX      Db 197 pIlqIhlypesdIhpqrtrrIInpIrlvDyvgvavIpravnpsrtravdltesIIrSa 256
XX
XX      120 HGCHTQYMANMMSVASIALAIWVGKDSKLGLVGVGHGCSPPRYVFPRLAYCE 173
XX      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX      Db 257 yInchIlykImmgvgastIsIkdg----hIvgIlaChnqIrkvlIpneItrkage 306
XX
XX      RESULT 6
XX      AAB26600 standard; peptide: 212 AA.

```

```

XX AAB26600;
AC
XX
XX 01-FEB-2001 (first entry)
DT
XX
XX Synechocystis sp phytochrome-related gene Cph2-N197 peptide.
DE
XX
XX Phytochrome; fluorescent apophytochrome; bilin; Cph.
KM
XX
XX Synechocystis sp.
OS
XX
XX WO200056355-A1.
PN
XX
XX 28-SEP-2000.
PD
XX
XX 14-MAR-2000; 2000MO-US06607.
PF
XX
XX 19-MAR-1999; 99US-0272809.
PR
XX
XX (REGC ) UNITV CALIFORNIA.
PA
XX
XX Lagarias JC;
PI
XX
XX WPI; 2000-602195/57.
DR
XX
XX Fluorescent apophytochrome-bilin conjugates, e.g. Cph2 (derived from
PT Synechocystis species) and phycoerythrobilin conjugate, useful as
PR fluorescent markers for biological research -
XX
XX Claim 5; Fig 3; 52pp; English.
PS
CC The present invention is related to fluorescent apophytochrome-bilin
CC conjugates, known as phytofluors. An apoprotein known as Cph2 from
CC Synechocystis species is used as the apophytochrome and the bilin is
CC preferably phycoerythrobilin. The phytofluors are useful as bilin
CC markers for phycoerythrobilin research. The phytofluors have a long wavelength
CC absorption maxima, a high molar absorption coefficient and the
CC recombinant apoproteins can spontaneously assemble with a variety of
CC bilin chromophore precursors. The present sequence is a phytochrome
CC related peptide from Synechocystis sp.
CC
XX
XX Sequence 212 AA;
SO
Query Match 18.6%; Score 190; DB 21; Length 212;
Best Local Similarity 31.5%; Pred. No. 2.6e-14;
Matches 51; Conservative 24; Mismatches 67; Indels 20; Gaps 5;
QY 5 RAISNLSQSLPGSGDIDGALCDTVYEDVQRLTGVGRWVWYQFHEPDHGKVSSE-IRRSLEPY 63
DB 20 raltirelt-----qviveeariflgvdvrxklyktaasgsevlaeavrraalpsl 70
QY 64 LGLHAPATDIPQARFLFKONFRNRMICDCNAPRVVVOSEEEK---RPLCLVNSTLRAPH 120
DB 71 lghlhpvedipqareeejngqfkmldavahrrk---shelsgrlsphehngnyttvd 127
QY 121 GCHTQYMANMGSVASLALAIYVVKGDSSKLMGLVGVGHHCSPR 162
DB 128 schligyllamgyvlssltvpm---qddqlwglwimavhnskpr 165
RESULT 7
AAB26593
ID AAB26593 standard; protein; 1276 AA.
XX
XX AAB26593;
XX
XX 01-FEB-2001 (first entry)
DT
XX
XX Synechocystis sp phytochrome-related gene Cph2.
DE
XX
XX Phytochrome; fluorescent apophytochrome; bilin; Cph.
XX

```



```

XX Synechocystis sp.
XX WO200056355-AI.
PN XX
PD 28-SEP-2000.
PE XX
PF 14-MAR-2000; 2000WO-US06607.
PR XX
PA (REGC ) UNIV CALIFORNIA.
XX
XX Lagarias JC;
PI
DR WPI: 2000-602195/57.
XX
XX Fluorescent apophytochrome-bilin conjugates, e.g. Cph2 (derived from
PT Synechocystis species) and phycoerythrobilin conjugate, useful as
PT fluorescent markers for biological research -
XX
PS Example 1; Page 45-46; 52pp; English.
XX
XX The present invention is related to Fluorescent apophytochrome-bilin
CC conjugates, known as phytofluors. An apoprotein known as Cph2 from
CC Synechocystis species is used as the apophytochrome and the bilin is
CC preferably phycoerythrobilin. The phytofluors are useful as fluorescent
CC markers for biological research. The phytofluors have a long wavelength
CC absorption maxima, a high molar absorptioin coefficient and the
CC recombinant apoproteins can spontaneously assemble with a variety of
CC bilin chromophore precursors. The present sequence is a phytochrome
CC related protein from Synechocystis sp.
XX
SQ Sequence 1276 AA;

Query Match          18.6%; Score 190; DB 21; Length 1276;
Best Local Similarity 31.5%; Pred. No. 3,5e-13;
Matches 51; Conservative 24; Mismatches 67; Indels 20; Gaps 5

OY 5 RAISRKQSLPGGDIGALCDTVVEVQRLTGYDRVVMYYOFHEDDHGEVSE-IKRSDLPEY 63
   *|||:::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB 20 raltrelt-----qvlveeariflgydrvtalktasdgsgevlaevnraalpsl 70
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

OY 64 IGLHPAIDIPQAAPFLFKQNRFMRICDCAATPVKVVOSEELK---RPLCLVNSTLRADH 120
   *|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB 71 Iqlhpviedipqaapeelngnfkmklavdaahrrkk--shelsgrlsptehsngnyltvd 127
   ||||::||||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

OY 121 GCHROYMANMGSVASLALAIVYKGDSKLWLLVGVGHCSPR 162
   ||||::||||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB 128 schigyllamgyllsltyvm---gdqqlwgimavhnskpr 165
   ||||::||||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

RESULT      8
AAB26597
DB D AAB26597 standard; protein: 844 AA.
AC
AA B26597;
AC
DT 01-FEB-2001 (first entry)
DE Synechocystis sp phytochrome-related gene Cph6.
XX
XX Phytochrome; fluorescent apophytochrome; bilin; Cph.
KW
XX
OS Synechocystis sp.
XX
XX WO200056355-AI.
PD
PD 28-SEP-2000.
PE
PE 14-MAR-2000; 2000WO-US06607.
PR
PR 19-MAR-1999; 99US-0272809.

```

```

XX      (REGC ) UNIV CALIFORNIA.
PA
XX      Lagarias JC;
PI
XX      WPI: 2000-602195/57.
DR
XX      Fluorescent apophytochrome-bilin conjugates, e.g. Cph2 (derived from
PT      Synecocystis species) and phycoerythrobilin conjugate, useful as
PR      fluorescent markers for biological research -
XX
PS      Example 1; Page 48-49; 52pp; English.
XX
CC      The present invention is related to fluorescent apophytochrome-bilin
CC      conjugates, known as phytofluors. An apoprotein known as Cph2 from
CC      Synecocystis species is used as the apophytochrome and the bilin is
CC      preferably phycoerythrobilin. The phytofluors are useful as fluorescent
CC      markers for biological research. The phytofluors have a long wavelength
CC      absorption maxima, a high molar absorbtion coefficient and the
CC      recombinant apoproteins can spontaneously assemble with a variety of
CC      bilin chromophore precursors. The present sequence is a phytochrome
CC      related protein from Synecocystis sp.
XX
SQ      Sequence      844 AA:
Query Match          12.0%; Score 123; DB 21; Length 844;
Best Local Similarity 22.2%; Pred. NO. 2.2e-05;
Matches 46; Conservative 39; Mismatches 72; Indels 50; Gaps 7.
OY      11 QSLPGDGLGALCDIYVEDYQRLTGIDRWAVYGFHEDDH---GEVVSSEIRSDLEPYLGHN 67
    ||| | : : :|||:::| |:::|| | : :| :
Db      472 gsl---dlptlfnvtvgelrflleadrvylfqfspdstdfsygnivaesvlabpfkplinsa 528
OY      68 YATDIPQAARFLFKONRYRMTCDCNAPFVKVVOSSSELKRPLCLYNSTLRAPHGCHTOYM 127
    | : | : :|||:::| |:::|| | : :| :
Db      529 leelfsmnyagryqgrtqviedlh-----gshlrq---chidfl 566
OY      128 ANMGSVASIALAIIVKKGSKSLMGLGVNGHCSPRYV-----PPFLRYA 171
    | : | : :|||:::| |:::|| | : :| :
Db      567 arlygranlvlpri----ndallkgldicbgcdssrvweqtetdillkqltngefelaigqa 622
OY      172 C--EFLMQAFGIQLDMELQLASQLAEK 196
    | : | : :|||:::| |:::|| | : :| :
Db      623 tlyeqagqelasknqlfvqltneleqk 649

RESULT   9
AAB26594 AAB26594 standard; protein; 481 AA.
XX ID AAB26594
XX AC AAB26594;
XX XX
DT      01-FEB-2001 (first entry)
XX XX
DE      Synecocystis sp phytochrome-related gene Cph3.
XX XX
KW      Phytochrome; fluorescent apophytochrome; bilin; Cph.
XX XX
OS      Synecocystis sp.
XX XX
PN      WO200056355-A1.
XX PD
XX      28-SEP-2000.
XX PF
XX      14-MAR-2000; 2000WO-US06607.
XX PR
XX      19-MAR-1999; 99US-0272809.
XX PA
XX      (REGC ) UNIV CALIFORNIA.
XX PI
XX      Lagarias JC;

```


THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OK protein - protein search, using sw model

Run on: June 7, 2002, 18:55:44 ; Search time 25.57 seconds
(without alignments)
187.228 Million cell updates/sec

Title: US-09-272-809-9
Perfect score: 1024
Sequence: 1 KIAVRAISRLQSLPGSDIGA.....QAFGLQLQMLQASLAERK 196

Scoring table: BL0SUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/CTUS.COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/backfilest.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	785	76.7	611	3 US-08-904-871-4	Sequence 4, Appli
2	785	76.7	1142	3 US-08-904-871-12	Sequence 12, Appli
3	733	71.6	554	3 US-08-904-871-1	Sequence 1, Appli
4	672	65.6	600	3 US-08-904-871-3	Sequence 3, Appli
5	663.5	64.8	600	3 US-08-904-871-2	Sequence 2, Appli
6	663.5	64.8	1129	3 US-08-904-871-11	Sequence 11, Appli
7	388	37.9	748	3 US-08-904-871-5	Sequence 5, Appli
8	388	37.9	748	3 US-08-904-871-6	Sequence 6, Appli
9	388	37.9	748	3 US-08-904-871-13	Sequence 13, Appli
10	72.5	7.1	353	1 US-08-197-792-43	Sequence 43, Appli
11	72.5	7.1	353	1 US-08-459-850-43	Sequence 43, Appli
12	72.5	7.1	353	1 US-08-459-214-43	Sequence 43, Appli
13	72.5	7.1	936	5 PCT-US93-05944-2	Sequence 6, Appli
14	71.5	7.0	5588	4 US-09-036-987A-6	Sequence 6, Appli
15	71.5	7.0	5588	4 US-09-370-700-6	Sequence 6, Appli
16	69	6.7	629	1 US-08-250-740-33	Sequence 33, Appli
17	69	6.7	629	1 US-07-695-42B-2	Sequence 2, Appli
18	69	6.7	1118	4 US-09-379-523-3	Sequence 3, Appli
19	68.5	6.7	2584	4 US-08-936-135-4	Sequence 4, Appli
20	67.5	6.6	308	4 US-09-347-803-8	Sequence 8, Appli
21	67.5	6.6	2588	3 US-08-936-135-2	Sequence 2, Appli
22	67	6.5	4928	4 US-09-036-987A-5	Sequence 5, Appli
23	67	6.5	4928	4 US-09-370-700-5	Sequence 5, Appli
24	66.5	6.5	349	1 US-08-197-792-33	Sequence 33, Appli
25	66.5	6.5	349	1 US-08-459-850-33	Sequence 33, Appli
26	66.5	6.5	349	1 US-08-459-214-33	Sequence 33, Appli
27	66.5	6.5	414	2 US-08-845-161A-2	Sequence 2, Appli

28	66.5	6.5	414	4 US-09-270-751-2	Sequence 2, Appli
29	65.5	6.4	178	4 US-09-147-928-4	Sequence 4, Appli
30	65.5	6.4	413	4 US-09-147-928-2	Sequence 2, Appli
31	65	6.3	206	4 US-09-311-311C-22	Sequence 22, Appli
32	65	6.3	313	4 US-09-347-803-25	Sequence 25, Appli
33	65	6.3	582	2 US-08-989-386-1	Sequence 1, Appli
34	64.5	6.3	399	2 US-08-742-621-4	Sequence 4, Appli
35	64.5	6.3	399	2 US-08-750-134A-5	Sequence 5, Appli
36	64.5	6.3	399	4 US-09-363-745-5	Sequence 5, Appli
37	64.5	6.3	993	1 US-08-468-557-2	Sequence 2, Appli
38	64	6.2	2548	4 US-09-172-422-1	Sequence 1, Appli
39	63.5	6.2	151	1 US-08-332-576-3	Sequence 3, Appli
40	63.5	6.2	151	5 PCT-US95-13672-3	Sequence 3, Appli
41	63.5	6.2	323	2 US-08-540-804-18	Sequence 18, Appli
42	63.5	6.2	323	2 US-08-218-265-18	Sequence 18, Appli
43	63.5	6.2	323	3 US-08-521-872-18	Sequence 18, Appli
44	63.5	6.2	323	4 US-08-590-399-18	Sequence 18, Appli
45	63.5	6.2	394	2 US-08-555-568B-17	Sequence 17, Appli

ALIGNMENTS

RESULT 1
US-08-904-871-4
; Sequence 4, Application US/08904871
; Patent No. 6046014
; GENERAL INFORMATION:
; APPLICANT: Lagarias, John C
; APPLICANT: Murphy, John T
; TITLE OF INVENTION: PHOTOLUORS AS FLUORESCENT LABELS
; FILE REFERENCE: 2500.134USO UC OTT Lagarias Patent
; CURRENT APPLICATION NUMBER: US/08/904,871
; CURRENT FILING DATE: 1997-08-01
; EARLIER APPLICATION NUMBER: 60/023,217
; EARLIER FILING DATE: 1996-08-02
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 611
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:: Sequence
US-08-904-871-4

Query Match 76.7%; Score 785; DB 3; Length 611;
Best Local Similarity 70.0%; Pred. No. 1.4e-90;
Matches 147; Conservative 20; Mismatches 29; Indels 14; Gaps 1;

QY 1 KIAVRAISRLQSLPGSDIGA...DVTVEVQRLTGYRVWVYGFHEDDGEVSEIRSDL 60
DB 203 KIAAARISRLQSLPGSDILLCDAVVEVRELGTDRVAVYFHEDEKEVLAIEIRSDL 262
QY 61 EBYLGLHYPATDIPQARLFKQNRVRLICDCAATPVKVVQSELRKRLCLVNSTLRAPH 120
DB 263 EBYLGLHYPATDIPQARLFKQNRVRLICDCAATPVKVVQSELRKRLCLVNSTLRAPH 322
QY 121 GCHTQYMANMGSVASLALAIIVKAGKDS-----KMKLVVGHHCSPRYVP 166
DB 323 GCHTQYMANMGSVASLALAIIVKAGKDS-----KMKLVVGHHCSPRYVP 382
QY 167 PLRYACEFLMAQFGLQLOMELQASLAERK 196
DB 383 PLRYACEFLMAQFGLQLOMELQASLAERK 412
RESULT 2
US-08-904-871-12
; Sequence 12, Application US/08904871
; Patent No. 6046014

```

; GENERAL INFORMATION:
; APPLICANT: Lagarias, John C
; APPLICANT: Murphy, John T
; TITLE OF INVENTION: PHYTOFLUORS AS FLUORESCENT LABELS
; FILE REFERENCE: 2500.134US0 UC OTT Lagarias Patent
; CURRENT APPLICATION NUMBER: US/08/904,871
; CURRENT FILING DATE: 1997-08-01
; EARLIER APPLICATION NUMBER: 60/023,217
; EARLIER FILING DATE: 1996-08-02
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO: 12
; LENGTH: 1142
; TYPE: PRT
; ORGANISM: Alga (Mesotetrahnum)
US-08-904-871-12
```

```

Query Match          76.7%; Score 785; DB 3; Length 1142;
Best Local Similarity 70.0%; Pred. No. 3,7e-90;
Matches 147; Conservative 20; Mismatches 29; Indels 14; Gaps 1;
```

```

Qy 1 KLAFAISRLQSLPGDICALCTVVEDVQRLTGYDRVAVVQFHEDDGEVVSSEIRSDL 60
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 203 KLAFAISRLQSLPGDICALCTVVEDVQRLTGYDRVAVVQFHEDDGEVVSSEIRSDL 262
Qy 61 EPLYGLHVPATDIPQARFLFKQNRVRLMCDNATPVVYVQSEELKRPCLVNSTLRAPH 120
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 263 EPLYGLHVPATDIPQARFLFKQNRVRLMCDNATPVVYVQSEELKRPCLVNSTLRAPH 322
Qy 121 GCHTOYMANMGVSASIALAIYVKGDS-----KLMGLVYGHHCSPRYVPF 166
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 333 GCHTOYMANMGVSASIALAIYVKGDS-----KLMGLVYGHHCSPRYVPF 382
Qy 167 PLRYACEFLMOAFGLQLOMELQASOLAER 196
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 383 PLRYACEFLMOAFGLQLOMELQASOLAER 412
```

```

RESULT 3
US-08-904-871-1
; Sequence 1, Application US/08904871
; Patent No. 6046014
; GENERAL INFORMATION:
; APPLICANT: Lagarias, John C
; APPLICANT: Murphy, John T
; TITLE OF INVENTION: PHYTOFLUORS AS FLUORESCENT LABELS
; FILE REFERENCE: 2500.134US0 UC OTT Lagarias Patent
; CURRENT APPLICATION NUMBER: US/08/904,871
; CURRENT FILING DATE: 1997-08-01
; EARLIER APPLICATION NUMBER: 60/023,217
; EARLIER FILING DATE: 1996-08-02
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO: 1
; LENGTH: 554
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:: Sequence
; OTHER INFORMATION: alignment fig. 6 Consensus
US-08-904-871-1
```

```

Query Match          71.6%; Score 733; DB 3; Length 554;
Best Local Similarity 72.5%; Pred. No. 4,8e-84;
Matches 145; Conservative 20; Mismatches 27; Indels 8; Gaps 5;
```

```

Qy 1 KLAFAISRLQSLPGDICALCTVVEDVQRLTGYDRVAVVQFHEDDGEVVSSEIRSDL 60
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 181 KLAFAISRLQSLPGDICALCTVVEDVQRLTGYDRVAVVQFHEDDGEVVSSEIRSDL 238
Qy 61 EPLYGLHVPATDIPQARFLFKQNRVRLMCDNATPVVYVQSEELKRPCLVNSTLRAPH 120
```

```

Db 239 EPLYGLHVPATDIPQARFLFKQNRVRLMCDNATPVVYVQSEELKRPCLVNSTLRAPH 296
Qy 121 GCHTOYMANMGVSASIALAIYVKGDS-----SKLMGLVYGHHCSPRYVPFPLRYACEFLM 176
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 297 GCHTOYMANMGVSASIALAIYVKGDS-----SKLMGLVYGHHCSPRYVPFPLRYACEFLM 356
Qy 177 QAFGLQLOMELQASOLAER 196
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 357 QAFGLQLOMELQASOLAER 376
```

```

RESULT 4
US-08-904-871-3
; Sequence 3, Application US/08904871
; Patent No. 6046014
; GENERAL INFORMATION:
; APPLICANT: Lagarias, John C
; APPLICANT: Murphy, John T
; TITLE OF INVENTION: PHYTOFLUORS AS FLUORESCENT LABELS
; FILE REFERENCE: 2500.134US0 UC OTT Lagarias Patent
; CURRENT APPLICATION NUMBER: US/08/904,871
; CURRENT FILING DATE: 1997-08-01
; EARLIER APPLICATION NUMBER: 60/023,217
; EARLIER FILING DATE: 1996-08-02
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO: 3
; LENGTH: 600
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:: Sequence
; OTHER INFORMATION: alignment fig. 6 Alpha
US-08-904-871-3
```

```

Query Match          65.6%; Score 672; DB 3; Length 600;
Best Local Similarity 60.5%; Pred. No. 2,9e-76;
Matches 127; Conservative 28; Mismatches 41; Indels 14; Gaps 1;
```

```

Qy 1 KLAFAISRLQSLPGDICALCTVVEDVQRLTGYDRVAVVQFHEDDGEVVSSEIRSDL 60
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 202 KLAFAISRLQSLPGDICALCTVVEDVQRLTGYDRVAVVQFHEDDGEVVSSEIRSDL 261
Qy 61 EPLYGLHVPATDIPQARFLFKQNRVRLMCDNATPVVYVQSEELKRPCLVNSTLRAPH 120
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 263 EPLYGLHVPATDIPQARFLFKQNRVRLMCDNATPVVYVQSEELKRPCLVNSTLRAPH 321
Qy 121 GCHTOYMANMGVSASIALAIYVKGDS-----SKLMGLVYGHHCSPRYVPF 166
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 322 SCHLOYMANMGVSASIALAIYVKGDS-----SKLMGLVYGHHCSPRYVPF 381
Qy 167 PLRYACEFLMOAFGLQLOMELQASOLAER 196
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 382 PLRYACEFLMOAFGLQLOMELQASOLAER 411
```

```

RESULT 5
US-08-904-871-2
; Sequence 2, Application US/08904871
; Patent No. 6046014
; GENERAL INFORMATION:
; APPLICANT: Lagarias, John C
; APPLICANT: Murphy, John T
; TITLE OF INVENTION: PHYTOFLUORS AS FLUORESCENT LABELS
; FILE REFERENCE: 2500.134US0 UC OTT Lagarias Patent
; CURRENT APPLICATION NUMBER: US/08/904,871
; CURRENT FILING DATE: 1997-08-01
; EARLIER APPLICATION NUMBER: 60/023,217
; EARLIER FILING DATE: 1996-08-02
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patent In Ver. 2.0
```



```
; SEQ ID NO 2
; LENGTH: 600
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Sequence
US-08-904-871-2
```

```
Query Match          64.8%; Score 663.5; DB 3; Length 600;
Best Local Similarity 59.7%; Pred. No. 3.4e-75;
Matches 126; Conservative 24; Mismatches 46; Indels 15; Gaps 1;
```

```
QY 1 KLAVAISRLOSIPGSDIGALCDTVEDVQRLTGYDRVMYQFHHDDHGEVSEIRSD 60
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 201 KLAARAIKISQSLPGSMSEVLCNTVYKEVFDLTGYDRVMYKFHHDDHGEVSEITKPEL 260
QY 61 EPLYGLHPATDIPQARFLFKQNRVIMICDGNATPVKVYVSEELKRPCLVNSTLRAPH 120
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 261 EPLYGLHPATDIPQARFLFKQNRVIMICDGNATPVKVYVSEELKRPCLVNSTLRAPH 320
QY 121 GCHTOYMANMGSVASLALAIYVKGKDSKLMGLVGHHCSPRYVP 165
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 321 SCHLOYMENMNSIASLVAVVYVNEDEDEASEQPAQOQKKKKLGLLVCHHESPRIYP 380
```

```
QY 166 PFLRYACEFLMQAFGLQLOMELQLASQLAEK 196
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 381 PFLRYACEFLMQAFGLQLOMELQLASQLAEK 411
```

```
RESULT 6
US-08-904-871-11
; Sequence 11, Application US/08904871
; Patent No. 6046014
; GENERAL INFORMATION:
; APPLICANT: Lagarias, John T
; TITLE OF INVENTION: PHYTOFLUORS AS FLUORESCENT LABELS
; FILE REFERENCE: 2500.134USO UC OTT Lagarias Patent
; CURRENT APPLICATION NUMBER: US/08/904,871
; EARLIER FILING DATE: 1997-08-01
; EARLIER APPLICATION NUMBER: 60/023,217
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 1129
; TYPE: PRT
; ORGANISM: Oat (Avena)
US-08-904-871-11
```

```
Query Match          64.8%; Score 663.5; DB 3; Length 1129;
Best Local Similarity 59.7%; Pred. No. 9.1e-75;
Matches 126; Conservative 24; Mismatches 46; Indels 15; Gaps 1;
```

```
QY 1 KLAVAISRLOSIPGSDIGALCDTVEDVQRLTGYDRVMYQFHHDDHGEVSEIRSD 60
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 201 KLAARAIKISQSLPGSMSEVLCNTVYKEVFDLTGYDRVMYKFHHDDHGEVSEITKPEL 260
QY 61 EPLYGLHPATDIPQARFLFKQNRVIMICDGNATPVKVYVSEELKRPCLVNSTLRAPH 120
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 261 EPLYGLHPATDIPQARFLFKQNRVIMICDGNATPVKVYVSEELKRPCLVNSTLRAPH 320
QY 121 GCHTOYMANMGSVASLALAIYVKGKDSKLMGLVGHHCSPRYVP 165
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 321 SCHLOYMENMNSIASLVAVVYVNEDEDEASEQPAQOQKKKKLGLLVCHHESPRIYP 380
QY 166 PFLRYACEFLMQAFGLQLOMELQLASQLAEK 196
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 381 PFLRYACEFLMQAFGLQLOMELQLASQLAEK 411
```

```
RESULT 7
US-08-904-871-5
; Sequence 5, Application US/08904871
; Patent No. 6046014
; GENERAL INFORMATION:
; APPLICANT: Lagarias, John T
; TITLE OF INVENTION: PHYTOFLUORS AS FLUORESCENT LABELS
; FILE REFERENCE: 2500.134USO UC OTT Lagarias Patent
; CURRENT APPLICATION NUMBER: US/08/904,871
; EARLIER FILING DATE: 1996-08-02
; EARLIER APPLICATION NUMBER: 60/023,217
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 528
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Sequence
US-08-904-871-5
```

```
Query Match          37.9%; Score 388; DB 3; Length 528;
Best Local Similarity 44.0%; Pred. No. 2.2e-40;
Matches 77; Conservative 37; Mismatches 53; Indels 8; Gaps 3;
```

```
QY 2 LAVRAISRLOSIPGSDIGALCDTVEDVQRLTGYDRVMYQFHHDDHGEVSEIRSD 61
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 139 MANAALNRLRQ--QANLRPEFDYIVAEVRMTGFEFRVMLRFDENNHGVIADKKDDME 196
QY 62 EPLYGLHPATDIPQARFLFKQNRVIMICDGNATPVKVYVSEELKRPCLVNSTLRAP 119
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 197 EPLYGLHPATDIPQARFLFKQNRVIMICDGNATPVKVYVSEELKRPCLVNSTLRAP 256
QY 120 HCHTOYMANMGSVASLALAIYVKGKDSKLMGLVGHHCSPRYVPFLRYACEF 174
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 257 YCHHLTYLKNMGVGSALFTSLIKDG---HLWGLIACHHQTPVIFPELRKACEF 307
```

```
RESULT 8
US-08-904-871-6
; Sequence 6, Application US/08904871
; Patent No. 6046014
; GENERAL INFORMATION:
; APPLICANT: Lagarias, John T
; TITLE OF INVENTION: PHYTOFLUORS AS FLUORESCENT LABELS
; FILE REFERENCE: 2500.134USO UC OTT Lagarias Patent
; CURRENT APPLICATION NUMBER: US/08/904,871
; EARLIER FILING DATE: 1997-08-01
; EARLIER APPLICATION NUMBER: 60/023,217
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 748
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Fig 10b seq
US-08-904-871-6
```

```
Query Match          37.9%; Score 388; DB 3; Length 748;
Best Local Similarity 44.0%; Pred. No. 3.8e-40;
Matches 77; Conservative 37; Mismatches 53; Indels 8; Gaps 3;
```

```
QY 2 LAVRAISRLOSIPGSDIGALCDTVEDVQRLTGYDRVMYQFHHDDHGEVSEIRSD 61
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

Db 139 MANAALNRLQ--QANLDFDYVIVEEVRMTGDFRVMILYRFDENNGHGVIAEDKRDDME 196
Oy 62 PYLGHTPATDIPQARLEFKONRYRMICDCN--ATPVKVOSEELKRLCLVNSTRAP 119
Db 197 PYLGHTPESDIPQARRLFHNPIRVIPDYGVAVPLTPAVNPSTNRVADLTESILRSA 256
Oy 120 HGHCHQVYMANNGSVASLALAIYVKGKSSKLMGLVYGHHSFRVYFPFLRYACEF 174
Db 257 YHCHLTLYLKNMGVASTLTSLIKDG----HLMGLIACHHQPVKVPIPFELKACEF 307

RESULT 9
US-08-904-871-13
Sequence 13, Application US/08904871
Patent No. 6046014
GENERAL INFORMATION:
APPLICANT: Lagarias, John C
APPLICANT: Murphy, John T
TITLE OF INVENTION: PHYTOFLUORS AS FLUORESCENT LABELS
FILE REFERENCE: 2500.134050 UC OTT Lagarias Patent
CURRENT APPLICATION NUMBER: US/08/904,871
CURRENT FILING DATE: 1997-08-01
EARLIER APPLICATION NUMBER: 60/023,217
EARLIER FILING DATE: 1996-08-02
NUMBER OF SEQ ID NOS: 16
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 13
LENGTH: 748
TYPE: PRT
ORGANISM: Unknown
FEATURE:
OTHER INFORMATION: Description of Unknown Organism: Cyanobacteria
US-08-904-871-13

Query Match 37.9%; Score 388; DB 3; Length 748;
Best Local Similarity 44.0%; Pred. No. 3.8e-40;
Matches 77; Conservative 37; Mismatches 53; Indels 8; Gaps 3;

Oy 2 LAVRISRLOSIPGGDICALCDTVVEDYORLTGYRVVYOPFHEDHGEVSEIRSDLE 61
Db 139 MANAALNRLQ--QANLDFDYVIVEEVRMTGDFRVMILYRFDENNGHGVIAEDKRDDME 196
Oy 62 PYLGHTPATDIPQARLEFKONRYRMICDCN--ATPVKVOSEELKRLCLVNSTRAP 119
Db 197 PYLGHTPESDIPQARRLFHNPIRVIPDYGVAVPLTPAVNPSTNRVADLTESILRSA 256
Oy 120 HGHCHQVYMANNGSVASLALAIYVKGKSSKLMGLVYGHHSFRVYFPFLRYACEF 174
Db 257 YHCHLTLYLKNMGVASTLTSLIKDG----HLMGLIACHHQPVKVPIPFELKACEF 307

RESULT 10
US-08-197-792-43
Sequence 43, Application US/08197792
Patent No. 5525488
GENERAL INFORMATION:
APPLICANT: Anthony J. Mason
APPLICANT: Peter H. Seeburg
TITLE OF INVENTION: Nucleic Acid Encoding the Alpha or Beta Chains of Inhibin and
NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/197,792
FILING DATE: 16-FEB-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/958414
FILING DATE: 08-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/744207
FILING DATE: 12-AUG-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/215466
FILING DATE: 05-JUL-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 06/906729
FILING DATE: 31-DEC-1986
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 06/827710
FILING DATE: 07-FEB-1986
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 06/783910
FILING DATE: 03-OCT-1985
ATTORNEY/AGENT INFORMATION:
NAME: Hasak, Janet E.
REGISTRATION NUMBER: 28,616
REFERENCE/DOCKET NUMBER: 297P2D4
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1896
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 43:
SEQUENCE CHARACTERISTICS:
LENGTH: 353 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-197-792-43

Query Match 7.1%; Score 72.5; DB 1; Length 353;
Best Local Similarity 34.4%; Pred. No. 1.1;
Matches 31; Conservative 14; Mismatches 30; Indels 15; Gaps 7;

Oy 37 RMYVQFHEHDEGEVSEI--RRSDLEPYLGHT--YPATDIPQARLEFKONRYRMICD--C 92
Db 144 RKVYV--FOEOGCDGMNVEKRVYDK--RSGWHTPFLTEAIOA---LFRGERRRLNDVQC 198
Oy 93 NA-----TPVKVVOSEELKRLCLVNSTL 116
Db 199 DSCQELAVPVFVDPGESHREPVVQARL 228

RESULT 11
US-08-459-850-43
Sequence 43, Application US/08459850
Patent No. 5663568
GENERAL INFORMATION:
APPLICANT: Anthony J. Mason
APPLICANT: Peter H. Seeburg
TITLE OF INVENTION: Nucleic Acid Encoding the Alpha or
TITLE OF INVENTION: Beta Chains of Inhibin and Method for Synthesizing Polypept
NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 kb floppy disk
COMPUTER: IBM PC compatible

```
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/459,850
FILING DATE: 02-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/197792
FILING DATE: 17-FEB-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/958414
FILING DATE: 08-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/744207
FILING DATE: 12-AUG-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/215466
FILING DATE: 05-JUL-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 06/906729
FILING DATE: 31-DEC-1986
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 06/827710
FILING DATE: 07-FEB-1986
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 06/783910
FILING DATE: 03-OCT-1985
ATTORNEY/AGENT INFORMATION:
NAME: Hasak, Janet E.
REGISTRATION NUMBER: 28,616
REFERENCE/DOCKET NUMBER: 297P2D5
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1896
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 43:
SEQUENCE CHARACTERISTICS:
LENGTH: 353 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-459-850-43

Query Match          7.1%; Score 72.5; DB 1; Length 353;
Best Local Similarity 34.4%; Pred. No. 1.1;
Matches 31; Conservative 14; Mismatches 30; Indels 15; Gaps 7;

QY 37 RVNYOFHEDHGEVSEI-RRSDLEPYLGH-YPATDIPQARFLFKONRVRMICD--C 92
    |||||  |||  :  :  :  |||  |  |  |  |||  |  |  |  |  |
Db 144 RVKYY-FOEGHGDGRMNVKRVDLK-RSGWHTFPLETAIQ---LFRGERRLMDVQC 198

QY 93 NA-----TPVKVVOSEELKRPLCLVNSTL 116
    :  :  ||  ||  ||  ||  :  :  |
Db 199 DSCQELAVPVFVDPGESHRPFVYVQARL 228

RESULT 12
US-08-459-214-43
; Sequence 43, Application US/08459214
; Patent No. 5716810
; GENERAL INFORMATION:
; APPLICANT: Anthony J. Mason
; APPLICANT: Peter H. Seeburg
; TITLE OF INVENTION: Nucleic Acid Encoding the Alpha or
; TITLE OF INVENTION: Beta Chains of Inhibin and Method for Synthesizing Polypeptide
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESS: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
```

```
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/459,214
FILING DATE: 02-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/197792
FILING DATE: 17-FEB-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/958414
FILING DATE: 08-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/744207
FILING DATE: 12-AUG-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/215466
FILING DATE: 05-JUL-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 06/906729
FILING DATE: 31-DEC-1986
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 06/827710
FILING DATE: 07-FEB-1986
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 06/783910
FILING DATE: 03-OCT-1985
ATTORNEY/AGENT INFORMATION:
NAME: Hasak, Janet E.
REGISTRATION NUMBER: 28,616
REFERENCE/DOCKET NUMBER: 297P2D6
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1896
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 43:
SEQUENCE CHARACTERISTICS:
LENGTH: 353 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-459-214-43

Query Match          7.1%; Score 72.5; DB 1; Length 353;
Best Local Similarity 34.4%; Pred. No. 1.1;
Matches 31; Conservative 14; Mismatches 30; Indels 15; Gaps 7;

QY 37 RVNYOFHEDHGEVSEI-RRSDLEPYLGH-YPATDIPQARFLFKONRVRMICD--C 92
    |||||  |||  :  :  :  |||  |  |  |  |||  |  |  |  |  |
Db 144 RVKYY-FOEGHGDGRMNVKRVDLK-RSGWHTFPLETAIQ---LFRGERRLMDVQC 198

QY 93 NA-----TPVKVVOSEELKRPLCLVNSTL 116
    :  :  ||  ||  ||  ||  :  :  |
Db 199 DSCQELAVPVFVDPGESHRPFVYVQARL 228

RESULT 13
PCT-US93-05944-2
; Sequence 2, Application PC/TUS9305944
; GENERAL INFORMATION:
; APPLICANT: Lin et al., Hun-Chi
; TITLE OF INVENTION: Molecular Cloning of the genes
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESS: Harris Brotman
; STREET: 401 B. St Ste 1700
; CITY: San Diego
; STATE: CA
```

COUNTRY: USA
ZIP: 92101-4297
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/05944
FILING DATE: 19930622
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Brozman, Harris F.
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 699-3630
TELEFAX: (619) 236-1048
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 936 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US93-05944-2

Query Match 7.0%; Score 72; DB 5; Length 936;
Best Local Similarity 22.1%; Pred. No. 5.7;
Matches 31; Conservative 19; Mismatches 62; Indels 28; Gaps 3;

QY 14 PGDIGALCDTVEDVQ---RLTGYDRVMYQFHHDDHGEVYSEIRRDLEPYLGLHYP 70
DB 711 PNDIKRANGPIVGGVTKGDLNGSDADTFYFDVKEGDVYIELPYGSSNFTWLVE 770
QY 71 TDIDQARFLFKONRVMICDCAATPVKVOSEELKRPCLVNSTLRAPHCCHTOYMANM 130
DB 771 GD-----DQNHIASGIDKNSKV-----GTFKATKGRHYFIYIKH 805
QY 131 GVSASLALAIYVKGKSSKL 150
DB 806 DSASNISYSLNKIGKNEKL 825

RESULT 14
US-09-036-987A-6
Sequence 6, Application US/09036987A
Patent No. 6143526
GENERAL INFORMATION:
APPLICANT: Baltz, Richard H.
APPLICANT: Broughton, Mary C.
APPLICANT: Crawford, Kathryn P.
APPLICANT: Madduri, Krishnamurthy
APPLICANT: Merlo, Donald J.
APPLICANT: Treadway, Patli J.
APPLICANT: Turner, Jan R.
APPLICANT: Waldron, Clive
TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dow Agrosciences LLC Patent Department
STREET: 9330 Zionsville Road
CITY: Indianapolis
STATE: Indiana
COUNTRY: USA
ZIP: 46268
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/036,987A
FILING DATE: 09-MAR-1998

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Stuart, Donald R.
REGISTRATION NUMBER: 28,479
REFERENCE/DOCKET NUMBER: 50,608
TELECOMMUNICATION INFORMATION:
TELEPHONE: (317)337-4816
TELEFAX: (317)337-4847
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 558 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-036-987A-6

Query Match 7.0%; Score 71.5; DB 4; Length 558;
Best Local Similarity 26.7%; Pred. No. 1e+02;
Matches 44; Conservative 20; Mismatches 64; Indels 37; Gaps 8;

QY 4 VRAISRLOSLPGDIGALCDTVEDVQRLTGYDR---VMYQFHHDDHGEVYSEIR-RSD 59
DB 3445 IRRNRASCTELADEGTLGVREHAAVLTGSSADYGVERRARDLGFDSLGYELRNR 3504
QY 60 LEPLYGLHYPATDI-----PQA-ARFLFKONRVMICDCAATPVKVOSEELKRPCLV 113
DB 3505 LAGVIGVRLPATFAVDYTPPALARFLHQE---LADIATTPAPV----- 3546
QY 114 STLRAPHCCHTOYMANMGSVASLALAIYVKGKSS--KLNGLYVG 156
DB 3547 TTRAP-----VAEDDLVAIVGMCREFPGOVSSPEELMLVAG 3584

RESULT 15
US-09-370-700-6
Sequence 6, Application US/09370700
Patent No. 6274350
GENERAL INFORMATION:
APPLICANT: Baltz, Richard H.
APPLICANT: Broughton, Mary C.
APPLICANT: Crawford, Kathryn P.
APPLICANT: Madduri, Krishnamurthy
APPLICANT: Treadway, Patli J.
APPLICANT: Turner, Jan R.
APPLICANT: Waldron, Clive
TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
FILE REFERENCE: 50489 DIV1
CURRENT APPLICATION NUMBER: US/09/370,700
CURRENT FILING DATE: 1999-08-09
EARLIER APPLICATION NUMBER: US 09/36987
NUMBER OF SEQ ID NOS: 39
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 6
LENGTH: 558
TYPE: PRT
ORGANISM: Saccharopolyspora spinosa
US-09-370-700-6

Query Match 7.0%; Score 71.5; DB 4; Length 558;
Best Local Similarity 26.7%; Pred. No. 1e+02;
Matches 44; Conservative 20; Mismatches 64; Indels 37; Gaps 8;

QY 4 VRAISRLOSLPGDIGALCDTVEDVQRLTGYDR---VMYQFHHDDHGEVYSEIR-RSD 59
DB 3445 IRRNRASCTELADEGTLGVREHAAVLTGSSADYGVERRARDLGFDSLGYELRNR 3504
QY 60 LEPLYGLHYPATDI-----PQA-ARFLFKONRVMICDCAATPVKVOSEELKRPCLV 113
DB 3505 LAGVIGVRLPATFAVDYTPPALARFLHQE---LADIATTPAPV----- 3546

Mon Jun 10 09:44:14 2002

us-09-272-809-9.ra1

Page 7

OY 114 STLRAHGHCHQYMANMGSVASIALAIYVKGDSS--KLMLVYG 156
:| | | : : : : :
Db 3547 TTTTAP-----VAEDDLVATVIGCGCRFPGOVSSPEELMLRLVAG 3584

Search completed: June 7, 2002, 18:55:46
Job time: 185 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 7, 2002, 18:56:32 ; Search time 37.51 Seconds
(without alignments)
502.093 million cell updates/sec

Title: US-09-272-809-9
Perfect score: 1024
Sequence: 1 KLAIVRAISRQSLPGDGIGA.....QAFGLQLQMLQLASQLAEK 196

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues
Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1024	100.0	1112	2	phytochrome E - Ar
2	840.5	82.1	1131	2	phytochrome - Scot
3	836	81.6	1135	2	phytochrome B - cu
4	828	80.9	1132	2	phytochrome B - co
5	806	78.7	1039	2	phytochrome B - so
6	802.5	78.4	1172	1	phytochrome B - Ar
7	794.5	77.6	1171	2	phytochrome B - ri
8	785	76.7	1142	2	phytochrome 1b - M
9	783	76.5	1164	2	phytochrome D - Ar
10	772	75.4	1156	2	phytochrome B - so
11	765.5	74.8	1136	2	phytochrome - Notw
12	763.5	74.6	577	2	phytochrome - moss
13	763.5	74.6	1129	2	phytochrome B - po
14	747	72.9	1134	2	phytochrome B - Mart
15	746	72.9	368	2	phytochrome - Doug
16	740.5	72.3	1132	2	phytochrome - moss
17	734.5	71.7	1135	2	phytochrome C - so
18	723.5	70.7	1303	1	phytochrome BI - S
19	717.5	70.1	211	1	phytochrome B2 - S
20	707	69.0	210	2	phytochrome C - Ar
21	706	68.9	1111	1	PHY3 protein - mai
22	687	67.1	1465	2	phytochrome A - Ar
23	672	65.6	1122	1	phytochrome A [imp
24	672	65.6	1122	2	phytochrome A - po
25	671.5	65.6	1123	2	phytochrome 3 - oa
26	669.5	65.4	1129	2	phytochrome 5 - oa
27	668.5	65.3	495	2	phytochrome A - pa
28	668	65.2	1129	2	phytochrome 4 - oa
29	666.5	65.1	1129	2	phytochrome 4 - oa

30	666	65.0	1124	1	FXPUZ	phytochrome - zucc
31	665.5	65.0	1124	2	S06856	phytochrome - gard
32	665.5	65.0	1128	2	S03728	phytochrome (clone
33	662.5	64.7	1125	2	T09835	phytochrome A - po
34	660.5	64.5	1131	2	J00382	phytochrome A - ma
35	646	63.1	1131	2	T07137	phytochrome A - so
36	584.5	57.1	190	2	S46926	phytochrome - Mats
37	561	54.8	156	2	T14837	phytochrome 2 - NO
38	548	53.5	189	2	S46927	phytochrome - Nym
39	521	50.9	197	2	S46928	phytochrome - Char
40	513	50.1	156	2	T14838	phytochrome 3 - No
41	507.5	49.6	314	2	T14865	phytochrome - Notw
42	501	48.9	156	2	T14839	phytochrome 4 - No
43	489	47.8	156	2	T14836	phytochrome - Notw
44	474	46.3	105	2	T17026	phytochrome E - ga
45	455.5	44.5	115	2	T09337	phytochrome - Doug

ALIGNMENTS

RESULT 1
S46313
phytochrome E - Arabidopsis thaliana
N:Alternate names: protein F1505.100
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 27-Jan-1995 #sequence_revision 27-Jan-1995 #text_change 05-May-2000
C:Accession: S46313; T14813; S41912
R:Clack, T.; Mathews, S.; Sharrock, R.A.
Plant Mol. Biol. 25, 413-427, 1994
A:Title: The phytochrome apoprotein family in Arabidopsis is encoded by five genes: t
A:Reference number: S46312; M01D:94325466
A:Accession: S46313
A:Molecule type: DNA
A:Residues: 1-1112 <CLN>
A:Cross-references: EMBL:76610; NID:9452815; PIDN:CA54075.1; PID:9452817
R:Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Mayer, K.
submitted to the Protein Sequence Database, August 1999
A:Reference number: Z18120
A:Accession: T14813
A:Molecule type: DNA
A:Residues: 1-1112 <BEV>
A:Cross-references: EMBL:AL10123; GSPDB:GN00062; ATSP:F1505.100
A:Experimental source: cultivar Columbia; BAC clone F1505
C:Genetics:
A:Gene: phyE; ATSP:F1505.100
A:Map position: 4
A:Introns: 665/1; 934/2; 1032/2
C:Superfamily: phytochrome; phytochrome homology
C:Keywords: chromoprotein; photoreceptor; phytochromobilin
F:63-569/Domain: phytochrome homology <PHY>
F:322/Binding site: phytochromobilin (Cys) (covalent) #status predicted

Query Match 100.0%; Score 1024; DB 2; Length 1112;
Best Local Similarity 100.0%; Pred. No. 1.8e-92;
Matches 196; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY	1	KLAIVRAISRQSLPGDGIGALCDTVVEDVQRLGTYRVVMVYQPHEDDHGEVSEIRRSDL	60
DB	201	KLAIVRAISRQSLPGDGIGALCDTVVEDVQRLGTYRVVMVYQPHEDDHGEVSEIRRSDL	260
OY	61	EPLYGLHYPATDIPQARLEFKONRVMICDCNATPVKVVQSEELKRPCLVNSTLRAPH	120
DB	261	EPLYGLHYPATDIPQARLEFKONRVMICDCNATPVKVVQSEELKRPCLVNSTLRAPH	320
OY	121	GCCTQYMANMGSAVALAIIVKGRKSSKIMGLVGHHSKPRVPPPLRACFFLWQARG	180
DB	321	GCCTQYMANMGSAVALAIIVKGRKSSKIMGLVGHHSKPRVPPPLRACFFLWQARG	380
OY	181	LQIMELQLASQLAEK	196
DB	381	LQIMELQLASQLAEK	396

RESULT 2

phytochrome - Scotch pine
C:Species: Pinus sylvestris (Scotch pine)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 20-Jun-2000
C:Accession: T09701
R:Wiegmann-Eirund, C.M.; Kolukisaoglu, H.U.
Submitted to the EMBL Data Library, March 1996
A:Reference number: Z16826
A:Accession: T09701
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1131 <WIE>
A:Cross-references: EMBL:X96738; NID:g1237083
A:Experimental source: isolate PSA 5.1
C:Superfamily: phytochrome; phytochrome homology
C:Keywords: chromoprotein; photoreceptor; phytochromobilin
F:75-587/Domain: phytochrome homology <PHY>
F:332/Binding site: phytochromobilin (Cys) (covalent) #status predicted

Query Match 82.1%; Score 840.5; DB 2; Length 1131;
Best Local Similarity 78.0%; Pred. No. 2.2e-74;
Matches 160; Conservative 15; Mismatches 21; Indels 9; Gaps 1;

QY 1 KLAIVRAISRLQSLPGDIGALCDTVEEDYQRLTGVDVWVYGFHDDHGEVYSEIRSDL 60
DB 211 KLAIVRAISRLQSLPGDVGLLCDTVEENRELTGIDRVWVYFHEDEHGEVVAESIRSDL 270
QY 61 EPLYGLHYPATDIPQARFLFKONRVKMDCAAPPVKVQSEELKRPCLVNSTLRAPH 120
DB 271 EPLYGLHYPATDIPQASRFLFKONRVKMDCAAPPVKVQSEELKRPCLVNSTLRAPH 330
QY 121 GCHTOYMANMGVSASIALAIIVYKGD-----SSKLMGLVVGHCSPRYVPPFLRYAC 171
DB 331 GCHAOYMANMGVSIRSLAMVVIINGNDEGGSGSRMSKMLGLVGHHTSPRAVPPFLRYA 390
QY 172 CEFLMQAFGLQLOMELQIASQLAEK 196
DB 391 CEFLMQAFGLQLMELQIASQLAEK 415

RESULT 3

phytochrome B - curled-leaved tobacco
C:Species: Nicotiana plumbaginifolia (curled-leaved tobacco)
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jul-2000
C:Accession: T16973
R:Hudson, M.E.; Robson, P.R.H.; Kraepiel, Y.; Caboche, M.; Smith, H.
Plant J. 12, 1091-2101, 1997
A:Title: Nicotiana plumbaginifolia hlg mutants have a mutation in a PHYB-type phytochrom
A:Reference number: Z18626; MUID:98079245
A:Accession: T16973
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1135 <HGD>
A:Cross-references: EMBL:Y14676; NID:g2370330; PIDN:CAA74992.1; PID:g2370331
A:Experimental source: cultivar PHND; Leaf
C:Genetics:
A:Gene: PHYB
C:Superfamily: phytochrome; phytochrome homology
C:Keywords: chromoprotein; photoreceptor; phytochromobilin
F:82-592/Domain: phytochrome homology <PHY>
F:338/Binding site: phytochromobilin (Cys) (covalent) #status predicted

Query Match 81.6%; Score 836; DB 2; Length 1135;
Best Local Similarity 77.0%; Pred. No. 6.2e-74;
Matches 157; Conservative 18; Mismatches 21; Indels 8; Gaps 1;

QX 1 KLAIVRAISRLQSLPGDIGALCDTVEEDYQRLTGIDRVWVYGFHDDHGEVYSEIRSDL 60

DB 217 KLAIVRAISRLQSLPGDVGKILCDTVEESVRELTDGVWVYFHEDEHGEVVAESIRPDL 276
QY 61 EPLYGLHYPATDIPQARFLFKONRVKMDCAAPPVKVQSEELKRPCLVNSTLRAPH 120
DB 277 EPLYGLHYPATDIPQASRFLFKONRVKMDCAAPPVKVQSEELKRPCLVNSTLRAPH 336
QY 121 GCHTOYMANMGVSASIALAIIVYKGD-----SSKLMGLVVGHCSPRYVPPFLRYAC 172
DB 337 GCHAOYMANMGVSIRSLAMVVIINGNDEAVGGRSMRLGLVGHHTSARCIPIFLRYAC 396
QY 173 CEFLMQAFGLQLOMELQIASQLAEK 196
DB 397 CEFLMQAFGLQLMELQIASQLAEK 420

RESULT 4

phytochrome B - common tobacco
C:Species: Nicotiana tabacum (common tobacco)
C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 21-Jan-2000
C:Accession: T03668; T03672
R:Kern, R.; Gasch, A.; Deak, M.; Kay, S.A.; Chua, N.H.
Plant Physiol. 102, 1363-1364, 1993
A:Title: PHYB of tobacco, a new member of the photoreceptor family.
A:Reference number: Z14996; MUID:94105358
A:Accession: T03668
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1132 <KER>
A:Cross-references: EMBL:L10114; NID:g295345; PIDN:AAA34092.1; PID:g295346
A:Experimental source: strain SR1; tissue-type etiolated seedling
R:Lopez-Juez, E.; Nagatani, A.; Tomizawa, K.; Deak, M.; Kern, R.; Kendrick, R.E.; Fur
Plant Cell 4, 241-251, 1992
A:Title: The cucumber long hypocotyl mutant lacks a light-stable PHYB-like phytochrom
A:Reference number: Z14997; MUID:92361250
A:Accession: T03672
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 457-506, 'S', 508-586, 'Q', 587-1132 <LOP>
A:Cross-references: EMBL:M65023; NID:g170286; PIDN:AAA34093.1; PID:g170287
C:Genetics:
A:Gene: PHYB
C:Superfamily: phytochrome; phytochrome homology
C:Keywords: chromoprotein; photoreceptor; phytochromobilin
F:80-589/Domain: phytochrome homology <PHY>
F:336/Binding site: phytochromobilin (Cys) (covalent) #status predicted

Query Match 80.9%; Score 828; DB 2; Length 1132;
Best Local Similarity 76.5%; Pred. No. 3.8e-73;
Matches 156; Conservative 18; Mismatches 22; Indels 8; Gaps 1;

QY 1 KLAIVRAISRLQSLPGDIGALCDTVEEDYQRLTGVDVWVYGFHDDHGEVYSEIRSDL 60
DB 215 KLAIVRAISRLQSLPGDVGKLLCDTVEESVRELTDGVWVYFHEDEHGEVVAESIRPDL 274
QY 61 EPLYGLHYPATDIPQARFLFKONRVKMDCAAPPVKVQSEELKRPCLVNSTLRAPH 120
DB 275 EPLYGLHYPATDIPQASRFLFKONRVKMDCAAPPVKVQSEELKRPCLVNSTLRAPH 334
QY 121 GCHTOYMANMGVSASIALAIIVYKGD-----SSKLMGLVVGHCSPRYVPPFLRYAC 172
DB 335 GCHAOYMANMGVSIRSLAMVVIINGNDEAVGGRSMRLGLVGHHTSARCIPIFLRYAC 394
QY 173 CEFLMQAFGLQLOMELQIASQLAEK 196
DB 395 CEFLMQAFGLQLMELQIASQLAEK 418

RESULT 5

phytochrome B - sorghum (fragment)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 7, 2002, 18:57:39 ; Search time 60.43 Seconds

(without alignments)
561.096 Million cell updates/sec

Title: US-09-272-809-9

Perfect score: 1024

Sequence: 1 KLAIVRAISRQSLPGSDIC.....QAFGLQLQMLEQLASLAEK 196

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: SPREMBL_19:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mammal:*
8: sp_mhc:*
9: sp_organelle:*
10: sp_phage:*
11: sp_plant:*
12: sp_rodent:*
13: sp_virus:*
14: sp_vertebrate:*
15: sp_unclassified:*
16: sp_virus:*
17: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	894	87.3	1137	10	Q9M6P6 Lycopersico
2	836.5	81.7	1130	10	Q24380 Solanum tub
3	836.5	81.7	1131	10	Q24380 Solanum tub
4	836	81.6	1135	10	Q24117 Lycopersico
5	828	80.9	1121	10	Q24117 Lycopersico
6	823	80.4	1151	10	Q9FQ03 Lycopersico
7	821	80.2	1151	10	Q9FQ03 Lycopersico
8	817	79.8	1151	10	Q9FQ04 Lycopersico
9	815	79.6	1151	10	Q9FQ04 Lycopersico
10	813	79.4	1151	10	Q9FQ04 Lycopersico
11	813	79.4	1151	10	Q9FQ04 Lycopersico
12	813	79.4	1151	10	Q9FQ04 Lycopersico
13	813	79.4	1151	10	Q9FQ04 Lycopersico
14	813	79.4	1151	10	Q9FQ04 Lycopersico
15	813	79.4	1151	10	Q9FQ04 Lycopersico
16	812.5	79.3	1151	10	Q9FQ04 Lycopersico

17	812	79.3	245	10	Q9FQW5	Q9FQW5 goupia glab
18	811	79.2	245	10	Q9FQW5	Q9FQW5 salacia imp
19	811	79.2	245	10	Q9FQW5	Q9FQW5 salacia imp
20	811	79.2	388	10	Q9FQW5	Q9FQW5 phragmites
21	810	79.1	245	10	Q9FQW5	Q9FQW5 morionia gr
22	809.5	79.1	392	10	Q9FQW5	Q9FQW5 lithacne p
23	809	79.0	361	10	Q9FQW5	Q9FQW5 periperygi
24	809	79.0	393	10	Q9FQW5	Q9FQW5 hakonechloa
25	806	78.7	245	10	Q9FQW5	Q9FQW5 eucryphia b
26	806	78.7	382	10	Q9FQW5	Q9FQW5 sorghum hal
27	806	78.7	1146	10	Q9FQW5	Q9FQW5 populus tri
28	805	78.6	245	10	Q9FQW5	Q9FQW5 aleurites m
29	805	78.6	393	10	Q9FQW5	Q9FQW5 dantonlops
30	805	78.6	393	10	Q9FQW5	Q9FQW5 panticum cap
31	804	78.5	245	10	Q9FQW5	Q9FQW5 gymnosporia
32	804	78.5	245	10	Q9FQW5	Q9FQW5 paxistima c
33	804	78.5	245	10	Q9FQW5	Q9FQW5 paxistima c
34	804	78.5	245	10	Q9FQW5	Q9FQW5 paxistima m
35	803.5	78.5	394	10	Q9FQW5	Q9FQW5 anemochloa
36	803.5	78.5	394	10	Q9FQW5	Q9FQW5 olira latif
37	803	78.4	245	10	Q9FQW5	Q9FQW5 gymnosporia
38	803	78.4	363	10	Q9FQW5	Q9FQW5 cuervea kap
39	803	78.4	365	10	Q9FQW5	Q9FQW5 loeseneriel
40	803	78.4	374	10	Q9FQW5	Q9FQW5 psammomya
41	802	78.3	245	10	Q9FQW5	Q9FQW5 puterlicki
42	802	78.3	245	10	Q9FQW5	Q9FQW5 atrostyax
43	802	78.3	360	10	Q9FQW5	Q9FQW5 wimmeria ac
44	802	78.3	374	10	Q9FQW5	Q9FQW5 cuervea int
45	802	78.3	374	10	Q9FQW5	Q9FQW5 reissantia

ALIGNMENTS

RESULT	ID	Q9M6P6	PRELIMINARY:	PRT: 1137 AA.
AC	Q9M6P6	01-OCT-2000 (TREMBLrel. 15, Created)		
DF	Q9M6P6	01-OCT-2000 (TREMBLrel. 15, Last sequence update)		
DT	Q9M6P6	01-DEC-2001 (TREMBLrel. 19, Last annotation update)		
DE	Q9M6P6	PHYTOCHROME E.		
GN	Q9M6P6	PHYE.		
OS	Q9M6P6	Lycopersicon esculentum (Tomato).		
OC	Q9M6P6	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Q9M6P6	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;		
OC	Q9M6P6	Asteridae; euasterids I; Solanales; Solanaceae; Solanum.		
OX	Q9M6P6	NCBI_TaxID=4081;		
RN	Q9M6P6	[1]		
RP	Q9M6P6	SEQUENCE FROM N.A.		
RX	Q9M6P6	MEDLINE-20188796; PubMed-10723737;		
RA	Q9M6P6	Alba R., Kelmenson P.M., Cordonnier-Pratt M.-M., Pratt L.H.;		
RT	Q9M6P6	"The phytochrome gene family in tomato and the rapid differential		
RT	Q9M6P6	evolution of this family in angiosperms.";		
RT	Q9M6P6	Mol. Biol. Evol. 17:362-373(2000).		
RL	Q9M6P6	EMBL, AF178571; AAP25812.1; -		
DR	Q9M6P6	InterPro: IPR003018; GAF.		
DR	Q9M6P6	InterPro: IPR003594; HATPase_c.		
DR	Q9M6P6	InterPro: IPR003661; His_Kin.		
DR	Q9M6P6	InterPro: IPR004359; His_Kin_slg.		
DR	Q9M6P6	InterPro: IPR000014; PAS		
DR	Q9M6P6	InterPro: IPR000700; PAS-assoc.C.		
DR	Q9M6P6	InterPro: IPR001294; Phytochrome.		
DR	Q9M6P6	Pfam: PF01590; GAF; 1.		
DR	Q9M6P6	Pfam: PF02518; HATPase_c; 1.		
DR	Q9M6P6	Pfam: PF00989; PAS; 2.		
DR	Q9M6P6	Pfam: PF00360; phytochrome; 1.		
DR	Q9M6P6	Pfam: PF00512; signal; 1.		
DR	Q9M6P6	PRINTS: PRO1033; PHYTOCHROME.		
DR	Q9M6P6	SMART: SMO0065; GAF; 1.		
DR	Q9M6P6	SMART: SMO0387; HATPase_c; 1.		
DR	Q9M6P6	SMART: SMO0388; HISKA; 1.		
DR	Q9M6P6	SMART: SMO0091; PAS; 2.		


```
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE PHYTOCHROME B (FRAGMENT).
GN PHB.
OC Goupia glabra.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids I; Celastraceae; Goupia.
OX NCBI_TaxID=39314;
RN [1]
RP SEQUENCE FROM N.A.
RA Simmons M.P., Clevinger C.C., Savolainen V., Archer R.H., Mathews S.,
RA Doyle J.J.;
RT "Phylogeny of the Celastraceae inferred from phytochrome B and
RT morphology."
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF216086; AAC49043.1; -.
DR InterPro: IPR003018; GAF.
DR InterPro: IPR001294; Phytochrome.
DR Pfam: PF01590; GAF; 1.
DR Pfam: PF00360; phytochrome; 1.
DR PRINTS: PRO1033; PHYTOCHROME.
DR SMART: SM00065; GAF; 1.
DR PROSITE: PS00245; PHYTOCHROME_1; 1.
DR PROSITE: PS50046; PHYTOCHROME_2; 1.
FT NON_TER 1
FT SEQUENCE 245 AA; 27109 MW; 55FE41CC97B30B CRC64;
SQ

Query Match 79.8%; Score 817; DB 10; Length 245;
Best Local Similarity 75.5%; Pred. No. 1.4e-76;
Matches 154; Conservative 21; Mismatches 21; Indels 8; Gaps 1;

OY 1 KLAVRAISRLOSIPGDIAGLCTDVEEDVORLTGYDRVWVYOFHEDHGEVSEIRSDL 60
DB 12 KLAVRAISRLOSIPGDIAGLCTDVEEDVORLTGYDRVWVYOFHEDHGEVSEIRSDL 71
OY 61 EPYGLHYPADIDPOARFELFKONRVKMICDCAIPVKVVOSEELKRPICLVNSTLRAPH 120
DB 72 EPYGLHYPADIDPOARFELFKONRVKMICDCAIPVKVVOSEELKRPICLVNSTLRAPH 131
OY 121 GCHTQYMANMGSVASLAIATVYKGD-----SSKIMGLVGHGCSPPRYVFFLRAC 172
DB 132 GCHTQYMANMGSVASLAIATVYKGD-----SSKIMGLVGHGCSPPRYVFFLRAC 191
OY 173 EFLMQAFGIQLMELQLASQLAEK 196
DB 192 EFLMQAFGIQLMELQLASQLAEK 215

RESULT 9
OYFQMO PRELIMINARY; PRT; 365 AA.
AC OYFQMO:
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE PHYTOCHROME B (FRAGMENT).
GN PHB.
OS Averrhoa carambola (Star fruit).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids I; Oxalidales; Oxalidaceae; Averrhoa.
OX NCBI_TaxID=28974;
RN [1]
RP SEQUENCE FROM N.A.
RA Simmons M.P., Clevinger C.C., Savolainen V., Archer R.H., Mathews S.,
RA Doyle J.J.;
RT "Phylogeny of the Celastraceae inferred from phytochrome B and
RT morphology."
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
```

```
DR EMBL: AF216092; AAC49049.1; -.
DR InterPro: IPR003017; DNA_Ligase.
DR InterPro: IPR003018; GAF.
DR InterPro: IPR001294; Phytochrome.
DR Pfam: PF01590; GAF; 1.
DR Pfam: PF00360; phytochrome; 1.
DR PRINTS: PRO1033; PHYTOCHROME.
DR SMART: SM00065; GAF; 1.
DR PROSITE: PS00697; DNA_LIGASE_A1; UNKNOWN_1.
DR PROSITE: PS50046; PHYTOCHROME_2; 1.
FT NON_TER 1
FT SEQUENCE 365 AA; 40359 MW; B464987BCE2E013C CRC64;
SQ

Query Match 79.8%; Score 817; DB 10; Length 365;
Best Local Similarity 75.5%; Pred. No. 2.3e-76;
Matches 154; Conservative 19; Mismatches 23; Indels 8; Gaps 1;

OY 1 KLAVRAISRLOSIPGDIAGLCTDVEEDVORLTGYDRVWVYOFHEDHGEVSEIRSDL 60
DB 12 KLAVRAISRLOSIPGDIAGLCTDVEEDVORLTGYDRVWVYOFHEDHGEVSEIRSDL 71
OY 61 EPYGLHYPADIDPOARFELFKONRVKMICDCAIPVKVVOSEELKRPICLVNSTLRAPH 120
DB 72 EPYGLHYPADIDPOARFELFKONRVKMICDCAIPVKVVOSEELKRPICLVNSTLRAPH 131
OY 121 GCHTQYMANMGSVASLAIATVYKGD-----SSKIMGLVGHGCSPPRYVFFLRAC 172
DB 132 GCHTQYMANMGSVASLAIATVYKGD-----SSKIMGLVGHGCSPPRYVFFLRAC 191
OY 173 EFLMQAFGIQLMELQLASQLAEK 196
DB 192 EFLMQAFGIQLMELQLASQLAEK 215

RESULT 10
OYXHB4 PRELIMINARY; PRT; 393 AA.
AC OYXHB4:
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE PHYTOCHROME B (FRAGMENT).
GN Sporobolus giganteus.
OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Chloridaceae; Eragrostidae; Sporobolus.
OX NCBI_TaxID=96053;
RN [1]
RP SEQUENCE FROM N.A.
RA Mathews S.Y., Tsai R.C., Kellogg E.A.;
RT "Phylogenetic structure in the grass family (Poaceae): evidence from
RT the nuclear gene phytochrome B."
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF137327; AAD41315.1; -.
DR InterPro: IPR000977; DNA_Ligase.
DR InterPro: IPR003018; GAF.
DR InterPro: IPR001294; Phytochrome.
DR Pfam: PF01590; GAF; 1.
DR Pfam: PF00360; phytochrome; 1.
DR PRINTS: PRO1033; PHYTOCHROME.
DR SMART: SM00065; GAF; 1.
DR PROSITE: PS00697; DNA_LIGASE_A1; UNKNOWN_1.
DR PROSITE: PS00245; PHYTOCHROME_1; 1.
DR PROSITE: PS50046; PHYTOCHROME_2; 1.
FT NON_TER 1
FT SEQUENCE 393 AA; 43566 MW; 119CABF4CFDD4B13 CRC64;
SQ

Query Match 79.6%; Score 815; DB 10; Length 393;
```


RN [1]
 RP SEQUENCE FROM N.A.
 RA Simmons M.P., Clevinger C.C., Savolainen V., Archer R.H., Mathews S.,
 RA Doyle J.J.:
 RT "Phylogeny of the Celastraceae inferred from phytochrome B and
 RT morphology."
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF216143; AAG49100.1; -.
 DR InterPro: IPR003018; GAF.
 DR InterPro: IPR001294; Phytochrome.
 DR Pfam: PF01590; GAF; 1.
 DR Pfam: PF00360; phytochrome; 1.
 DR PRINTS: PR01033; PHYTOCHROME.
 DR SMART: SM00065; GAF; 1.
 DR PROSITE: PS00245; PHYTOCHROME_1; 1.
 DR PROSITE: PS50046; PHYTOCHROME_2; 1.
 FT NON_TER 1 1
 FT SEQUENCE 374 AA; 41226 MW; BF2B27A8B6BD8BC CRC64;

Query Match 79.4%; Score 813; DB 10; Length 374;
 Best Local Similarity 74.5%; Pred. No. 6.3e-76;
 Matches 152; Conservative 23; Mismatches 21; Indels 8; Gaps 1;

OY 1 KLAVRAISRLOSLPGSDIGALCDTVVEDYQRLTGYDRVWVYGFHEDHGEVSEIRSD 60
 DB 12 KLAVRAISRLOSLPGSDIKLDCDAVENVRNLTYGDRVWVYGFHEDHGEVSEIRSD 71
 OY 61 EPYLGHTPATDIPQARLFKONRYRMICDCAATPVKVOSEELKRPCLYNSTLRAPH 120
 DB 72 EPYLGHTPATDIPQASRFLFKONRYRMIVDCHSTPVSYIQQDGLMQPLCLVGSSTLRAPH 131
 OY 121 GCHTOYMANNGSVASLALAIYVKGK-----SSKLGVLVGHHCSPRYVFPPLRYAC 172
 DB 132 GCHAOYMANNGSVASLALAIYVKGK-----SSKLGVLVGHHCSPRYVFPPLRYAC 191
 OY 173 EFLMQAFGLQLOMELQASOLAELK 196
 DB 192 EFLMQAFGLQLOMELQASOLAELK 215

RESULT 14
 O9FOP1 PRELIMINARY; PRT; 374 AA.
 AC O9FOP1;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE PHYTOCHROME B (FRAGMENT).
 GN PHYB.
 OS Salicopsis glomerata.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids I; Celastraceae; Salicopsis.
 OX NCBI_TaxID=123472;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Simmons M.P., Clevinger C.C., Savolainen V., Archer R.H., Mathews S.,
 RA Doyle J.J.:
 RT "Phylogeny of the Celastraceae inferred from phytochrome B and
 RT morphology."
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF216145; AAG49122.1; -.
 DR InterPro: IPR003018; GAF.
 DR InterPro: IPR001294; Phytochrome.
 DR Pfam: PF01590; GAF; 1.
 DR Pfam: PF00360; phytochrome; 1.
 DR PRINTS: PR01033; PHYTOCHROME.
 DR SMART: SM00065; GAF; 1.
 DR PROSITE: PS00245; PHYTOCHROME_1; 1.
 DR PROSITE: PS50046; PHYTOCHROME_2; 1.
 FT NON_TER 1 1

FT NON_TER 374 374
 SQ SEQUENCE 374 AA; 41178 MW; DB7E21E29E493AB8 CRC64;

Query Match 79.4%; Score 813; DB 10; Length 374;
 Best Local Similarity 75.0%; Pred. No. 6.3e-76;
 Matches 153; Conservative 23; Mismatches 20; Indels 8; Gaps 1;

OY 1 KLAVRAISRLOSLPGSDIGALCDTVVEDYQRLTGYDRVWVYGFHEDHGEVSEIRSD 60
 DB 12 KLAVRAISRLOSLPGSDIKLDCDAVENVRNLTYGDRVWVYGFHEDHGEVSEIRSD 71
 OY 61 EPYLGHTPATDIPQARLFKONRYRMICDCAATPVKVOSEELKRPCLYNSTLRAPH 120
 DB 72 EPYLGHTPATDIPQASRFLFKONRYRMIVDCHSTPVSYIQQDGLMQPLCLVGSSTLRAPH 131
 OY 121 GCHTOYMANNGSVASLALAIYVKGK-----SSKLGVLVGHHCSPRYVFPPLRYAC 172
 DB 132 GCHAOYMANNGSVASLALAIYVKGK-----SSKLGVLVGHHCSPRYVFPPLRYAC 191
 OY 173 EFLMQAFGLQLOMELQASOLAELK 196
 DB 192 EFLMQAFGLQLOMELQASOLAELK 215

RESULT 15
 O9XH99 PRELIMINARY; PRT; 388 AA.
 AC O9XH99;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE PHYTOCHROME B (FRAGMENT).
 GN PHYB.
 OS Molinia caerulea (moor grass).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
 OC Arundinoideae; Arundineae; Molinia.
 OX NCBI_TaxID=38689;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Mathews S.Y., Tsai R.C., Kellogg E.A.:
 RT "Phylogenetic structure in the grass family (Poaceae): evidence from
 RT the nuclear gene phytochrome B."
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF137312; AAD41300.1; -.
 DR InterPro: IPR000977; DNA_ligase.
 DR InterPro: IPR003018; GAF.
 DR InterPro: IPR001294; Phytochrome.
 DR Pfam: PF01590; GAF; 1.
 DR Pfam: PF00360; phytochrome; 1.
 DR PRINTS: PR01033; PHYTOCHROME.
 DR SMART: SM00065; GAF; 1.
 DR PROSITE: PS00697; DNA_ligase; 1.
 DR PROSITE: PS00245; PHYTOCHROME_1; 1.
 DR PROSITE: PS50046; PHYTOCHROME_2; 1.
 FT NON_TER 1 1
 FT SEQUENCE 388 AA; 42910 MW; BEADB2ECDB08E286 CRC64;

Query Match 79.4%; Score 813; DB 10; Length 388;
 Best Local Similarity 74.0%; Pred. No. 6.6e-76;
 Matches 154; Conservative 23; Mismatches 19; Indels 12; Gaps 2;

OY 1 KLAVRAISRLOSLPGSDIGALCDTVVEDYQRLTGYDRVWVYGFHEDHGEVSEIRSD 60
 DB 15 KLAVRAISRLOSLPGSDIKLDCDAVENVRNLTYGDRVWVYGFHEDHGEVSEIRSD 74
 OY 61 EPYLGHTPATDIPQARLFKONRYRMICDCAATPVKVOSEELKRPCLYNSTLRAPH 120
 DB 75 EPYLGHTPATDIPQASRFLFKONRYRMIVDCHSTPVSYIQQDGLMQPLCLVGSSTLRAPH 134

Mon Jun 10 09:44:16 2002

us-09-272-809-9.rspt

Page 8

```

Oy 121 GCHHGYNNMMSVSLSLAIIV-----KKQSS--KLMLVYGHHCSPRYVEPL 168
    |||||.....:|||.....:|||||..|||:|
Db 135 GCHQGYNNMMSIASLVAVIIIIIGGDEGTARSGISASAKIMGLVYCHHTSPRLHPL 194
    |||||.....:|||||.....:|||||..|||:|
Oy 169 RYACEFLMAGFLOLMETOLASLAKR 196
    |||||.....:|||||.....:|||||..|||:|
Db 195 RYACEFLMAGFLOLMETOLASLAKR 222
    |||||.....:|||||.....:|||||..|||:|

```

Search completed: June 7, 2002, 18:57:40
Job time: 239 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 7, 2002, 18:58:02 ; Search time 19.14 Seconds
(without alignments)
396.501 Million cell updates/sec

Title: US-09-272-809-9
Sequence: 1 KIAA0191RLQSLPGSDIGA.....QAFGLDQLMELASOLAEX 196

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1024	100.0	1112	1	P42498 arabidopsis
2	914	89.3	1115	1	P55004 parhbtis n
3	840.5	82.1	1131	1	P41046 pinus sylve
4	828	80.9	1132	1	P28130 nicotiana t
5	806	78.7	1178	1	P39527 sorghum bic
6	802.5	78.4	1172	1	P14713 arabidopsis
7	794.5	77.6	1171	1	P25764 oryza sativ
8	783	76.5	1164	1	P42497 arabidopsis
9	772	75.4	1156	1	P42499 glycine max
10	765.5	74.8	1136	1	Q40762 picea abies
11	763.5	74.6	1121	1	Q39557 ceratodon p
12	763.5	74.6	1129	1	P34094 solanum tub
13	758.5	74.1	1118	1	P42496 adiantum ca
14	747	72.9	1134	1	O01505 selaginella
15	740.5	72.3	1132	1	P38505 physcomitre
16	734.5	71.7	1135	1	P39528 sorghum bic
17	732.5	70.7	1137	1	Q95419 oryza sativ
18	723.5	70.7	1307	1	P25848 ceratodon p
19	722.5	70.6	1124	1	P33529 mengoecia s
20	706	68.9	1111	1	P14714 arabidopsis
21	672	65.6	1122	1	P14712 arabidopsis
22	671.5	65.6	1123	1	P33733 solanum tub
23	671.5	65.6	1124	1	P33530 nicotiana t
24	668.5	65.3	494	1	P06595 avena sativ
25	668.5	65.3	1131	1	P39526 sorghum bic
26	668	65.2	1129	1	P55141 petroselinu
27	666.5	65.1	1128	1	P06594 avena sativ
28	666	65.0	1124	1	P06592 cucurbita p
29	665.5	65.0	1124	1	P39573 lathyrus sa
30	665.5	65.0	1124	1	P15001 pisum sativ
31	663.5	64.8	1128	1	P10931 oryza sativ
32	663.5	64.8	1128	1	P06593 avena sativ
33	662.5	64.7	1125	1	O49934 populus tre

34	660.5	64.5	1131	1	PHYA_MAIZE	P19862 zea mays (m
35	646	63.1	1131	1	PHYA_SOYBN	P42500 glycine max
36	388	37.9	741	1	PHY1_SYNY3	Q05168 synchocyst
37	379	37.0	765	1	PHYA_ANASP	O91cc2 anabaena sp
38	366.5	35.8	728	1	BPHY_PSEAE	O9huc3 pseudomonas
39	356	34.8	751	1	PHYB_ANASP	O9r6x3 anabaena sp
40	315	30.8	755	1	BPHY_DEIRA	O9r2a4 delinococcus
41	190	18.6	1276	1	PHY2_SYNY3	O55434 synchocyst
42	85.5	8.3	633	1	SIR_SYNY3	P72854 synchocyst
43	79	7.7	333	1	ETPA_RAT	P13803 rattus norv
44	79	7.7	989	1	RPOC_LEUME	P44892 leuconostoc
45	76	7.4	192	1	CCA_AERHY	P45749 aeromonas h

ALIGNMENTS

RESULT 1
ID PHYE_ARATH STANDARD: PRT; 1112 AA.
AC P42498;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Phytochrome E.
GN PHYE OR AT4G18130 OR F1J5J5.100.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. LANDSBERG ERRECTA;
RX MEDLINE=94325466; PubMed=8049367;
RA "Clack T., Mathews S., Sharrock R.A.;
RT The phytochrome apoprotein family in Arabidopsis is encoded by five
RL genes: the sequences and expression of PHYD and PHYE.";
RL Plant Mol. Biol. 25:413-427(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=20083486; PubMed=10617198;
RA Meyer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,
RA Pohl T., Duesterhoeft A., Stiekema W., Entlan K.-D., Terryn N.,
RA Harris B., Ansoorge W., Brandt P., Grivell L.A., Rieger M.,
RA Weichselgartner M., de Simone V., Obermaier B., Meche R., Mueller M.,
RA Kreis M., Delseny M., Pridmore P., Watson M., Schmidheini T.,
RA Reichert B., Portetelle D., Perez-Alonso M., Bouly M., Bancroft I.,
RA Vos P., Heijzel J., Zimmermann W., Wedler H., Ridley P.,
RA Langham S.-A., McCullagh B., Billam L., Robben J.,
RA Van der Schueren J., Grymonprez B., Chung Y.-J., Vandenbussche F.,
RA Breken M., Weltjens I., Voet M., Bastiens I., Aert R., Deltor E.,
RA Weltzengerger T., Boche G., Ransperger U., Hilbert H., Braun M.,
RA Holzman P., Brandt A., Peters S., Van Slaveren M., Ditske W.,
RA Moeller J., Klein Lankhorst R., Rose M., Haut J., Koetter P.,
RA Berner S., Hempel S., Feldpausch M., Lamberth S., Van den Deele H.,
RA De Keyser A., Buysaert C., Gielen J., Villarroel R., De Clercq R.,
RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McElroy K., Mayes R.,
RA Pettett A., Rajandream M.A., Lyne M., Benes V., Reckmann S.,
RA Borikova D., Bloeker H., Scharfe M., Grimm M., Loehner T.-H.,
RA Dose S., de Haan M., Maier A.C., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Farthmann B., Grandpré K., Dauner D., Herzi A.,
RA Neumann S., Argitlon A., Vitale D., Liguori R., Pivrandi E.,
RA Massenet O., Ougley F., Clabaud G., Mendel A., Reiber R.,
RA Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
RA Chedot F., Cooke R., Berger C., Montfort A., Casacuberta S.,
RA Gibbons T., Weber N., Vandenbol M., Barges M., Terol J., Torres A.,
RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Belke C.,
RA Feilman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
RA Zaccaria P., Beyer M., Wilson R.K., de la Bastide M., Habermann K.,

RA Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
 RA Sekhon M., Murray J., Sheet P., Cordes M., Abu-Thelaidh J.,
 RA Stenking T., Rajicki J., Graves T., Harmon G., Edwards J.,
 RA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
 RA Mix P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
 RA Kramer J., Fulton L., Mardis E., Dance M., Pepin K., Hillier L.,
 RA Nelson J., Speith J., Ryan E., Andrews S., Geisel C., Layman D.,
 RA Du H., All J., Berghoff A., Jones K., Drone K., Cotton M., Joshua C.,
 RA Antonolu B., Zidanic M., Strong C., Sun H., Lamar B., Jordan C.,
 RA Ma P., Zhong J., Preston R., Vill D., Shekhar M., Matero A., Shah R.,
 RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,
 RA Granet S., Shohdy N., Hasegawa A., Hamed A., Lodhi M., Johnson A.,
 RA Chen E., Marra M., Martensen R., McComdie W.R.,
 RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis
 thaliana.";
 RL Nature 402:769-777(1999).
 CC -1- FUNCTION: REGULATOR PHOTORECEPTOR WHICH EXISTS IN TWO FORMS THAT
 ARE REVERSIBLY INTERCONVERTIBLE BY LIGHT: THE PR FORM THAT ABSORBS
 MAXIMALLY IN THE RED REGION OF THE SPECTRUM AND THE PFR FORM THAT
 ABSORBS MAXIMALLY IN THE FAR-RED REGION. PHOTOCOVERSION OF PR IN
 PFR INDUCES AN ARRAY OF MORPHOGENIC RESPONSES, WHEREAS
 RECONVERSION OF PFR TO PR CANCELS THE INDUCTION OF THOSE
 RESPONSES. PFR CONTROLS THE EXPRESSION OF A NUMBER OF NUCLEAR
 GENES, INCLUDING THOSE ENCODING THE SMALL SUBUNIT OF RIBULOSE-
 BISPHOSPHATE CARBOXYLASE, CHLOROPHYLL A/B BINDING PROTEIN,
 PROTOCHLOROPHYLLIDE REDUCTASE, RNA, ETC. IT ALSO CONTROLS
 THE EXPRESSION OF ITS OWN GENE(S) IN A NEGATIVE FEEDBACK FASHION.
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- PTM: CONTAINS ONE COVALENTLY LINKED TETRAPYRROLE CHROMOPHORE.
 CC -1- SIMILARITY: BELONGS TO THE PHYTOCHROME FAMILY.
 CC -1- SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 HISTIDINE KINASE DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X76610; CAA54075.1; -;
 DR EMBL: AL101023; CAA53654.1; -;
 DR EMBL: AL161548; CAB78815.1; -;
 DR Mendel; 7190; ARACH.PHYE.1;
 DR InterPro: IPR003018; GAF.
 DR InterPro: IPR003594; HATPase_C.
 DR InterPro: IPR004359; HIS_KIN_sig.
 DR InterPro: IPR003661; HIS_kina.
 DR InterPro: IPR000014; PAS.
 DR InterPro: IPR001294; Phytochrome.
 DR Pfam: PF01590; GAF; 1.
 DR Pfam: PF02518; HATPase_C; 1.
 DR Pfam: PF00989; PAS; 2.
 DR Pfam: PF00360; phytochrome; 1.
 DR Pfam: PF00512; signal; 1.
 DR PRINTS: PR01033; PHYTOCHROME.
 DR SMART: SM00065; GAF; 1.
 DR SMART: SM00387; HATPase_C; 1.
 DR SMART: SM00388; HSKA; 1.
 DR SMART: SM00091; PAS; 2.
 DR PROSITE: PS50109; HIS_KIN; 1.
 DR PROSITE: PS50112; PAS; 2.
 DR PROSITE: PS00245; PHYTOCHROME_1; 1.
 DR PROSITE: PS50046; PHYTOCHROME_2; 1.
 KW Transcription regulation; Photoreceptor; Phytochrome; Chromophore;
 KW Repeat: Multigene family.
 FT DOMAIN 595 666 PAS 1.
 FT DOMAIN 732 803 PAS 2.
 FT DOMAIN 877 1096 HISTIDINE KINASE.
 FT BINDING 322 322 CHROMOPHORE (BY SIMILARITY).
 SQ SEQUENCE 1112 AA; 122587 MW; 4C235B83F6D9DA28 CRC64;

Query Match 100.0%; Score 1024; DB 1; Length 1112;
 Best Local Similarity 100.0%; Pred. No. 6.6e-96;
 Matches 196; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 KLAIVRAISRLSGPGDIGALCDTVYEDVQRLTGDRVAVYGFHEDDHGEVSEIRSDL 60
 DB 201 KLAIVRAISRLSGIPGDIGALCDTVYEDVQRLTGDRVAVYGFHEDDHGEVSEIRSDL 260
 QY 61 EYLGLHVPATIPQARFLFKONVRMTCCDNATPVKVYVSEELKRPCLVNSTLRAPH 120
 DB 261 EYLGLHVPATIPQARFLFKONVRMTCCDNATPVKVYVSEELKRPCLVNSTLRAPH 320
 QY 121 GCHTYMANMGSVASIALAIYVKGRDSSKLMGLVYGHHCSPRYVPFPLRYACEFLMQAFG 180
 DB 321 GCHTYMANMGSVASIALAIYVKGRDSSKLMGLVYGHHCSPRYVPFPLRYACEFLMQAFG 380
 QY 181 LQLOMELQASLAEK 196
 DB 381 LQLOMELQASLAEK 396
 RESULT 2
 ID PHYE_PHANT STANDARD; PRT; 1115 AA.
 AC P55004;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DE 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Phytochrome E.
 GN PHYE.
 OS Pharbitis nil (Viollet) (Japanese morning glory).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; easterids I; Solanales; Convolvulaceae; Ipomoea.
 NCBI_TaxID=35983;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP TISSUE=Seedling cotyledon;
 RA Zheng C.C., O'Neill S.D.;
 RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: REGULATOR PHOTORECEPTOR WHICH EXISTS IN TWO FORMS THAT
 ARE REVERSIBLY INTERCONVERTIBLE BY LIGHT: THE PR FORM THAT ABSORBS
 MAXIMALLY IN THE RED REGION OF THE SPECTRUM AND THE PFR FORM THAT
 ABSORBS MAXIMALLY IN THE FAR-RED REGION. PHOTOCOVERSION OF PR IN
 PFR INDUCES AN ARRAY OF MORPHOGENIC RESPONSES, WHEREAS
 RECONVERSION OF PFR TO PR CANCELS THE INDUCTION OF THOSE
 RESPONSES. PFR CONTROLS THE EXPRESSION OF A NUMBER OF NUCLEAR
 GENES, INCLUDING THOSE ENCODING THE SMALL SUBUNIT OF RIBULOSE-
 BISPHOSPHATE CARBOXYLASE, CHLOROPHYLL A/B BINDING PROTEIN,
 PROTOCHLOROPHYLLIDE REDUCTASE, RNA, ETC. IT ALSO CONTROLS
 THE EXPRESSION OF ITS OWN GENE(S) IN A NEGATIVE FEEDBACK FASHION.
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- PTM: CONTAINS ONE COVALENTLY LINKED TETRAPYRROLE CHROMOPHORE.
 CC -1- SIMILARITY: BELONGS TO THE PHYTOCHROME FAMILY.
 CC -1- SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 HISTIDINE KINASE DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U39787; AAA84970.1; -;
 DR InterPro: IPR003018; GAF.
 DR InterPro: IPR003594; HATPase_C.
 DR InterPro: IPR004359; HIS_KIN_sig.
 DR InterPro: IPR003661; HIS_kina.
 DR InterPro: IPR000700; PAS-assoc_C.

DR InterPro: IPR000014; PAS.
 DR InterPro: IPR001294; Phytochrome.
 DR Pfam: PF01590; GAF; 1.
 DR Pfam: PF02518; HATPase_c; 1.
 DR Pfam: PF00989; PAS; 2.
 DR Pfam: PF00360; phytochrome; 1.
 DR Pfam: PF00512; signal; 1.
 DR PRINTS: PRO1033; PHYTOCHROME.
 DR SMART: SM00065; GAF; 1.
 DR SMART: SM00387; HATPase_c; 1.
 DR SMART: SM00091; HIS_KIN; 1.
 DR SMART: SM00981; PAS; 2.
 DR PROSITE: PS50109; HIS_KIN; 1.
 DR PROSITE: PS50112; PAC; 1.
 DR PROSITE: PS50112; PAS; 2.
 DR PROSITE: PS00245; PHYTOCHROME_1; 1.
 DR PROSITE: PS50046; PHYTOCHROME_2; 1.
 KW Transcription regulation; Photoreceptor; Phytochrome; Chromophore;
 KM Repeat.
 FT DOMAIN 598 669 PAS 1.
 FT DOMAIN 672 728 PAC.
 FT DOMAIN 732 803 PAS 2.
 FT DOMAIN 880 1100 HISTIDINE KINASE.
 FT BINDING 318 318 CHROMOPHORE (BY SIMILARITY).
 SQ SEQUENCE 1115 AA; 124328 MW; 081A4154E147800 CRC64;

Query Match 89.3%; Score 914; DB 1; Length 1115;
 Best Local Similarity 87.2%; Pred. No. 9.7e-85;
 Matches 171; Conservative 12; Mismatches 13; Indels 0; Gaps 0;

QY 1 KLAVAIRISLQSLPGDICALDVTYEDYQRLTGDRVWVYGFHEDHGEVSEIRSL 60
 DB 197 KLAVAIRISLQSLPGDICALDVTYEDYQRLTGDRVWVYGFHEDHGEVSEIRSL 256
 QY 61 EPYGLHYPATIDIPQARFLFKONRYRMICDCNATPVKVOSEELKRPCLVNSTLRAP 120
 DB 257 EPYGLHYPATIDIPQARFLFKONRYRMICDCNATPVKVOSEELKRPCLVNSTLRAP 316
 QY 121 GCHTOYMANMGSVASLALAIIVKRGDKSLMGLVGVGHSCSPRYVPEPLRYAEFLMQA 180
 DB 317 GCHTOYMANMGSVASLALAIIVKRGDKSLMGLVGVGHSCSPRYVPEPLRYAEFLMQA 376
 QY 181 LQLOMELQASLAQLEK 196
 DB 377 LQLYMELQASLAQLEK 392

RESULT 3

PHY_PINSY STANDARD; PRT; 1131 AA.
 ID PHY_PINSY STANDARD; PRT; 1131 AA.
 AC 041046;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Phytochrome.
 OS Pinus sylvestris (Scots pine).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus.
 OX NCBI_TaxID=3349;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PSA 5.1;
 RA Wiegman-Eirund C.M., Kolukisaoglu H.U.;
 RL Submitted (MAR-1996) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: REGULATORY PHOTORECEPTOR WHICH EXISTS IN TWO FORMS THAT
 ARE REVERSIBLY INTERCONVERTIBLE BY LIGHT. THE PR FORM THAT ABSORBS
 MAXIMALLY IN THE RED REGION OF THE SPECTRUM AND THE PFR FORM THAT
 ABSORBS MAXIMALLY IN THE FAR-RED REGION. PHOTOCOVERSION OF PR IN
 PFR INDUCES AN ARRAY OF MORPHOGENIC RESPONSES, WHEREAS
 RECONVERSION OF PFR TO PR CANCELS THE INDUCTION OF THOSE
 RESPONSES. PFR CONTROLS THE EXPRESSION OF A NUMBER OF NUCLEAR
 CC GENES INCLUDING THOSE ENCODING THE SMALL SUBUNIT OF RIBULOSE-

CC BISPHOSPHATE CARBOXYLASE, CHLOROPHYLL A/B BINDING PROTEIN,
 CC PROTOCHLOROPHYLLIDE REDUCTASE, RNA, ETC. IT ALSO CONTROLS
 CC THE EXPRESSION OF ITS OWN GENE(S) IN A NEGATIVE FEEDBACK FASHION.
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
 CC -1- PFM: CONTAINS ONE COVALENTLY LINKED TETRAPYRROLE CHROMOPHORE.
 CC -1- SIMILARITY: BELONGS TO THE PHYTOCHROME FAMILY.
 CC -1- SIMILARITY: CONTAINS 2 PAS (PER-ANT-SIM) DIMERIZATION DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 HISTIDINE KINASE DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-slb.ch/announce/>
 CC or send an email to license@isb-slb.ch).
 CC -----
 CC EMBL: X96738; CAA65510.1; -.
 DR InterPro: IPR003018; GAF.
 DR InterPro: IPR003594; HATPase_c.
 DR InterPro: IPR004359; HIS_KIN_sig.
 DR InterPro: IPR003661; HIS_KIN.
 DR InterPro: IPR001610; PAC.
 DR InterPro: IPR000014; PAS.
 DR InterPro: IPR001294; Phytochrome.
 DR Pfam: PF01590; GAF; 1.
 DR Pfam: PF02518; HATPase_c; 1.
 DR Pfam: PF00989; PAS; 2.
 DR Pfam: PF00360; phytochrome; 1.
 DR Pfam: PF00512; signal; 1.
 DR PRINTS: PRO1033; PHYTOCHROME.
 DR SMART: SM00065; GAF; 1.
 DR SMART: SM00387; HATPase_c; 1.
 DR SMART: SM00086; HIS_KIN; 1.
 DR SMART: SM00091; PAS; 2.
 DR PROSITE: PS50109; HIS_KIN; 1.
 DR PROSITE: PS50112; PAC; 2.
 DR PROSITE: PS00245; PHYTOCHROME_1; 1.
 DR PROSITE: PS50046; PHYTOCHROME_2; 1.
 KW Transcription regulation; Photoreceptor; Phytochrome; Chromophore;
 KM Repeat.
 FT DOMAIN 621 692 PAS 1.
 FT DOMAIN 755 826 PAS 2.
 FT DOMAIN 903 1123 HISTIDINE KINASE.
 FT BINDING 332 332 CHROMOPHORE (BY SIMILARITY).
 SQ SEQUENCE 1131 AA; 126254 MW; D63A2008FA9862EB CRC64;

Query Match 82.1%; Score 840.5; DB 1; Length 1131;
 Best Local Similarity 78.0%; Pred. No. 2.9e-77;
 Matches 160; Conservative 15; Mismatches 21; Indels 9; Gaps 1;

QY 1 KLAVAIRISLQSLPGDICALDVTYEDYQRLTGDRVWVYGFHEDHGEVSEIRSL 60
 DB 211 KLAVAIRISLQSLPGDICALDVTYEDYQRLTGDRVWVYGFHEDHGEVSEIRSL 270
 QY 61 EPYGLHYPATIDIPQARFLFKONRYRMICDCNATPVKVOSEELKRPCLVNSTLRAP 120
 DB 271 EPYGLHYPATIDIPQARFLFKONRYRMICDCNATPVKVOSEELKRPCLVNSTLRAP 330
 QY 121 GCHTOYMANMGSVASLALAIIVKRGDKSLMGLVGVGHSCSPRYVPEPLRYA 171
 DB 331 GCHTOYMANMGSVASLALAIIVKRGDKSLMGLVGVGHSCSPRYVPEPLRYA 390
 QY 172 CEFLMQAQLQLOMELQASLAQLEK 196
 DB 391 CEFLMQAQLQLOMELQASLAQLEK 415

RESULT 4
 PHYB_TOBAC

ID PHYB_TOBAC STANDARD: PRT: 1132 AA.
AC P29130;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Phytochrome B.
GN PHYB.
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids II; Solanales; Solanaceae; Nicotiana.
OC NCBI_TaxID=4097;
RX MEDLINE=94105358; PubMed=8278560;
RA Kern R., Gasch A., Deak M., Kay S.A., Chua N.H.;
RT "PhyB of tobacco, a new member of the phytochrome family.";
RL Plant Physiol. 102:1363-1364(1993).
RN [2]
RP SEQUENCE OF 457-1132 FROM N.A.
RX MEDLINE=92361250; PubMed=1498594;
RA Lopez-Juez E., Nagatani A., Tomizawa K.I., Deak M., Kern R.,
RA Kendrick R.E., Furuya M.;
RT "The cucumber long hypocotyl mutant lacks a light-stable PHYB-like
phytochrome.";
RL Plant Cell 4:241-251(1992).
CC -1- FUNCTION: REGULATORY PHOTORECEPTOR WHICH EXISTS IN TWO FORMS THAT
ARE REVERSIBLY INTERCONVERTIBLE BY LIGHT: THE PR FORM THAT ABSORBS
MAXIMALLY IN THE RED REGION OF THE SPECTRUM AND THE PFR FORM THAT
ABSORBS MAXIMALLY IN THE FAR-RED REGION. PHOTOCONVERSION OF PR IN
PFR INDUCES AN ARRAY OF MORPHOGENIC RESPONSES, WHEREAS
RECONVERSION OF PFR TO PR CANCELS THE INDUCTION OF THOSE
RESPONSES. PFR CONTROLS THE EXPRESSION OF A NUMBER OF NUCLEAR
GENES INCLUDING THOSE ENCODING THE SMALL SUBUNIT OF RIBULOSE-
BISPHOSPHATE CARBOXYLASE, CHLOROPHYLL A/B BINDING PROTEIN,
PHOTODIODEPHOSPHORYLATION, ETC. IT ALSO CONTROLS
THE EXPRESSION OF ITS OWN GENE(S) IN A NEGATIVE FEEDBACK FASHION.
CC -1- SUBUNIT: HOMODIMER.
CC -1- PFM: CONTAINS ONE COVALENTLY LINKED TETRAPYRROLE CHROMOPHORE.
CC -1- SIMILARITY: BELONGS TO THE PHYTOCHROME FAMILY.
CC -1- SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 HISTIDINE KINASE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L10114; AAA34092.1; -;
DR EMBL: M65023; AAA34093.1; -;
DR Mendei: 1321; NICta:PHYB.1.
DR InterPro: IPR003018; GAF.
DR InterPro: IPR003594; HATPase_C.
DR InterPro: IPR004359; HIS_KIN_sig.
DR InterPro: IPR003661; HIS_KINa.
DR InterPro: IPR001610; PAC.
DR InterPro: IPR000014; PAS.
DR InterPro: IPR001294; Phytochrome.
DR Pfam: PF01590; GAF; 1.
DR Pfam: PF02518; HATPase_C; 1.
DR Pfam: PF00989; PAS; 2.
DR Pfam: PF00360; phytochrome; 1.
DR PRINTS: PF00512; signal; 1.
DR PRINTS: PR01033; PHYTOCHROME.
DR SMART: SM00065; GAF; 1.
DR SMART: SM00387; HATPase_C; 1.
DR SMART: SM00388; HIS_KA; 1.
DR SMART: SM00086; PAC; 1.
DR SMART: SM00091; PAS; 2.

DR PROSITE: PS50109; HIS_KIN; 1.
DR PROSITE: PS50112; PAS; 2.
DR PROSITE: PS00245; PHYTOCHROME_1; 1.
DR PROSITE: PS50046; PHYTOCHROME_2; 1.
KW Transcription regulation; Photoreceptor; Phytochrome; Chromophore;
KW Repeat; Multigene family.
FT DOMAIN 623 694 PAS 1.
FT DOMAIN 757 828 PAS 2.
FT DOMAIN 905 1125 HISTIDINE KINASE.
FT BINDING 336 336 CHROMOPHORE (BY SIMILARITY).
FT CONFLICT 507 507 L -> S (IN REF. 2).
FT CONFLICT 586 586 L -> LQ (IN REF. 2).
SQ SEQUENCE 1132 AA; 125809 MW; 457F09C024C0F608 CRC64;

Query Match 80.9%; Score 828; DB 1; Length 1132;
Best local similarity 76.5%; Pred. No. 5,4e-76;
Matches 156; Conservative 18; Mismatches 22; Indels 8; Gaps 1;

QY 1 KLAVERAISRLQSLPGDIGALCDTVEDVQRLTGYDRVWVYQFHEDDGEVVSIRSDL 60
DB 215 KLAVERAISRLQSLPGDVKLDCDTVESVRELTYGDRVWVYKFHDEHGEVVAESKIPDL 274
QY 61 EYILGHPATDIPQARPLFKQNRVMTCDQNAIPYKVSSEELKRPCLYNSTLRAPH 120
DB 275 EYILGHPATDIPQASRPLFKQNRVMTVDCHAPPRVQDESIMQPLCLGVTLRAPH 334
QY 121 GCHTGYMANMGVSAVLATLAIYVKGD-----SSKLGLVGHGHSPPRYPPPLRYAC 172
DB 335 GCHADYMANMGSIATLITAVITINGNDEAVGRSSRLMGLVGHGHSARCLIPPLRYAC 394
QY 173 EFLMQAFGLQLOMELQSLAEK 196
DB 395 EFLMQAFGLQLMELQSLAEK 418

RESULT 5
PHYB_SORBI STANDARD: PRT: 1178 AA.
AC P93527;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Phytochrome B.
GN PHYB OR MA3.
OS Sorghum bicolor (Sorghum) (Sorghum vulgare).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Panicoideae; Andropogoneae; Sorghum.
OC NCBI_TaxID=4558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. 58M;
RX MEDLINE=20188796; PubMed=10723737;
RA Alpha R., Kelmenson P.M., Cordonier-Pratt M.-M., Pratt L.H.;
RT "The phytochrome gene family in tomato and the rapid differential
evolution of this family in angiosperms.";
RL Mol. Biol. Evol. 17:362-373(2000).
RN [2]
RP SEQUENCE OF 208-1178 FROM N.A.
RC STRAIN-CV. 58M;
RX MEDLINE=97198556; PubMed=9046599;
RA Childs K.L., Miller F.R., Cordonier-Pratt M.-M., Pratt L.H.,
RA Morgan P.W., Muller J.E.;
RT "The Sorghum bicolor photoperiod sensitivity gene, Ma3, encodes a
phytochrome B.";
RL Plant Physiol. 113:611-619(1997).
CC -1- FUNCTION: REGULATORY PHOTORECEPTOR WHICH EXISTS IN TWO FORMS THAT
ARE REVERSIBLY INTERCONVERTIBLE BY LIGHT: THE PR FORM THAT ABSORBS
MAXIMALLY IN THE RED REGION OF THE SPECTRUM AND THE PFR FORM THAT
ABSORBS MAXIMALLY IN THE FAR-RED REGION. PHOTOCONVERSION OF PR IN
PFR INDUCES AN ARRAY OF MORPHOGENIC RESPONSES, WHEREAS
RECONVERSION OF PFR TO PR CANCELS THE INDUCTION OF THOSE


```

DR EMBL: L09362; -: NOT_ANNOTATED_CDS.
DR EMBL: AC005724; AAD08948.1; -.
DR PIR: S07718; FKMOB.
DR PIR: JQ2141; JQ2141.
DR Mendel: 1310; Arabid. Phyl. 1.
DR InterPro: IPR003018; GAF.
DR InterPro: IPR003594; HATPase_C.
DR InterPro: IPR004359; HIS_KIN_S19.
DR InterPro: IPR003661; HIS_KIN_A.
DR InterPro: IPR000014; PAS.
DR InterPro: IPR001294; Phytochrome.
DR Pfam: PF01590; GAF; 1.
DR Pfam: PF02518; HATPase_C; 1.
DR Pfam: PF00989; PAS; 2.
DR Pfam: PF00360; Phytochrome; 1.
DR Pfam: PF00512; signal; 1.
DR PRINTS: PR01033; PHYTOCHROME.
DR SMART: SM00065; GAF; 1.
DR SMART: SM00387; HATPase_C; 1.
DR SMART: SM00388; HSKA; 1.
DR SMART: SM00091; PAS; 2.
DR PROSITE: PS50109; HIS_KIN; 1.
DR PROSITE: PS50112; PAS; 2.
DR PROSITE: PS00245; PHYTOCHROME_1; 1.
DR PROSITE: PS50046; PHYTOCHROME_2; 1.
DR Transcription regulation; Photoreceptor; Phytochrome; Chromophore;
KW Repeat; Multigene family.
FT DOMAIN 652 723 PAS 1.
FT DOMAIN 786 857 PAS 2.
FT DOMAIN 934 1153 HISTIDINE KINASE.
FT BINDING 357 357 GLY/SER-RICH.
FT BINDING 1172 AA; 129331 MW; 7B5348CB1091B813 CRC64;
SQ SEQUENCE 1172 AA; 129331 MW; 7B5348CB1091B813 CRC64;

Query Match 78.4%; Score 802.5; DB 1; Length 1172;
Best Local Similarity 72.9%; Pred. No. 2.2e-73;
Matches 151; Conservative 23; Mismatches 22; Indels 11; Gaps 1;

QY 1 KLAVRAISRLOSLPGDIGALCDIVVEYQRLTGDRMYVQFHEDDHGEVSEIRSDL 60
DB 236 KLAVRAISRLOSLPGDIGALCDIVVEYQRLTGDRMYVQFHEDDHGEVSEIRSDL 295
QY 61 EPTVGLHPATDIPQARFLFKONRVRMICDNPVKVVOSEELKRLCLVNSTLRAPH 120
DB 296 EPTVGLHPATDIPQARFLFKONRVRMICDNPVKVVOSEELKRLCLVNSTLRAPH 355
QY 121 GCHTOYANMGSVASLALAIYV-----KGKDSKLMGLVGHHCSPRYVPPFLR 169
DB 356 GCHTOYANMGSVASLALAIYV-----KGKDSKLMGLVGHHCSPRYVPPFLR 415
QY 170 YACEFLMQAEGLOLQMLQSLAELK 196
DB 416 YACEFLMQAEGLOLQMLQSLAELK 442

RESULT 7
PRTB_ORYSA STANDARD; PRT; 1171 AA.
AC P25764;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Phytochrome B.
GN PHVB OR PHYL.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_taxid=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. INDICA-IR36; TISSUE=Seedling shoot;

```

```

RX MEDLINE-91172131; PubMed-2005872;
RA Dehesh K., Tepperman J., Christensen A.H., Quail P.H.;
RT "PhyB is evolutionarily conserved and constitutively expressed in
RT rice seedling shoots."
RL Mol. Gen. Genet. 225:305-313(1991).
CC -!- FUNCTION: REGULATORY PHOTORECEPTOR WHICH EXISTS IN TWO FORMS THAT
CC ARE REVERSIBLY INTERCONVERTIBLE BY LIGHT: THE PR FORM THAT ABSORBS
CC MAXIMALLY IN THE RED REGION OF THE SPECTRUM AND THE PFR FORM THAT
CC ABSORBS MAXIMALLY IN THE FAR-RED REGION. PHOTOCOVERSION OF PR IN
CC PFR INDUCES AN ARRAY OF MORPHOGENIC RESPONSES, WHEREAS
CC RECONVERSION OF PFR TO PR CANCELS THE INDUCTION OF THOSE
CC RESPONSES. PFR CONTROLS THE EXPRESSION OF A NUMBER OF NUCLEAR
CC GENES INCLUDING THOSE ENCODING THE SMALL SUBUNIT OF RIBULOSE-
CC BISPHOSPHATE CARBOXYLASE, CHLOROPHYLL A/B BINDING PROTEIN,
CC PROTOCHLOROPHYLLIDE REDUCTASE, RNA, ETC. IT ALSO CONTROLS
CC THE EXPRESSION OF ITS OWN GENE(S) IN A NEGATIVE FEEDBACK FASHION.
CC -!- SUBUNIT: HOMODIMER.
CC -!- PFM: CONTAINS ONE COVALENTLY LINKED TETRAPPTROLE CHROMOPHORE.
CC -!- SIMILARITY: BELONGS TO THE PHYTOCHROME FAMILY.
CC -!- SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 HISTIDINE KINASE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X57563; CAA40795.2; -.
DR PIR: S14065; S14065.
DR PIR: S14065; S14065.
DR Mendel: 1319; Oryza: Phyl. 1.
DR InterPro: IPR003018; GAF.
DR InterPro: IPR003594; HATPase_C.
DR InterPro: IPR004359; HIS_KIN_S19.
DR InterPro: IPR003661; HIS_KIN_A.
DR InterPro: IPR000014; PAS.
DR InterPro: IPR001294; Phytochrome.
DR Pfam: PF01590; GAF; 1.
DR Pfam: PF02518; HATPase_C; 1.
DR Pfam: PF00989; PAS; 2.
DR Pfam: PF00360; Phytochrome; 1.
DR Pfam: PF00512; signal; 1.
DR PRINTS: PR01033; PHYTOCHROME.
DR SMART: SM00387; HATPase_C; 1.
DR SMART: SM00388; HSKA; 1.
DR SMART: SM00091; PAS; 2.
DR PROSITE: PS50109; HIS_KIN; 1.
DR PROSITE: PS50112; PAS; 2.
DR PROSITE: PS00245; PHYTOCHROME_1; 1.
DR PROSITE: PS50046; PHYTOCHROME_2; 1.
KW Transcription regulation; Photoreceptor; Phytochrome; Chromophore;
KW Repeat; Multigene family.
FT DOMAIN 661 732 PAS 1.
FT DOMAIN 795 866 PAS 2.
FT DOMAIN 943 1161 HISTIDINE KINASE.
FT DOMAIN 39 51 POLY-GLY.
FT BINDING 364 364 CHROMOPHORE (BY SIMILARITY).
FT BINDING 1171 AA; 128384 MW; E8292E88B769BF16 CRC64;
SQ SEQUENCE 1171 AA; 128384 MW; E8292E88B769BF16 CRC64;

Query Match 77.6%; Score 794.5; DB 1; Length 1171;
Best Local Similarity 71.8%; Pred. No. 1.4e-72;
Matches 150; Conservative 23; Mismatches 23; Indels 13; Gaps 2;

QY 1 KLAVRAISRLOSLPGDIGALCDIVVEYQRLTGDRMYVQFHEDDHGEVSEIRSDL 60
DB 243 KLAVRAISRLOSLPGDIGALCDIVVEYQRLTGDRMYVQFHEDDHGEVSEIRSDL 302
QY 61 EPTVGLHPATDIPQARFLFKONRVRMICDNPVKVVOSEELKRLCLVNSTLRAPH 120
DB 243 KLAVRAISRLOSLPGDIGALCDIVVEYQRLTGDRMYVQFHEDDHGEVSEIRSDL 302

```


KW Repeat. 629 699 PAS 1.
FT DOMAIN 762 833 PAS 2.
FT DOMAIN 913 1132 HISTIDINE KINASE.
FT BINDING 336 336 CHROMOPHORE (BY SIMILARITY).
SQ SEQUENCE 1136 AA; 126066 MW; 074AE01498453E93 CRC64;

Query Match 74.8%; Score 765.5; DB 1; Length 1136;
Best Local Similarity 68.9%; Pred. No. 1.2e-69;
Matches 144; Conservative 23; Mismatches 29; Indels 13; Gaps 2;

QY 1 KLAARAIKRLQSLPGDGLCDYVEVYQRLTGIDRVWVYFHEDDHGEVSEIRSL 60
DB 215 KLAARAIKRLQSLPGDGLCDYVEVYQRLTGIDRVWVYFHEDDHGEVSEIRSL 274
QY 61 EPIGLHYPATIDIPQARFLFKONRYRMICDCAFPVAVIYDKRLQSLGSLTRAPH 120
DB 275 EPIGLHYPATIDIPQARFLFKONRYRMICDCAFPVAVIYDKRLQSLGSLTRAPH 334
QY 121 GCHTYMANMGSVASIALAIYV--KGKDS-----KLMGLVGHGHSPPRYVPEP 167
DB 335 GCHTYMANMGSVASIALAIYV--KGKDS-----KLMGLVGHGHSPPRYVPEP 394
QY 168 LRYACEFLMQAFGLQLOMELQSLAEK 196
DB 395 LRYACEFLMQAFGLQLOMELQSLAEK 423

RESULT 11
PHY2_CERPU STANDARD; PRT; 1121 AA.
AC Q39557;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Phytochrome 2.
GN PH2.
OS Ceratodon purpureus (Moss).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;
OC Bryophyta; Dicranidae; Dicranales; Ditrichaceae; Ceratodon.
ON NCBI_TaxID=3225;
RN [1]
RP SEQUENCE FROM N.A.
RA Lamparter T., Miltmann F.;
RT "GRP;PHY2, a 'normal' phytochrome in Ceratodon.";
RL (In) Plant Gene Register PGR96-067.
CC -1- FUNCTION: REGULATORY PHOTORECEPTOR WHICH EXISTS IN TWO FORMS THAT ARE REVERSIBLY INTERCONVERTIBLE BY LIGHT. THE PR FORM THAT ABSORBS MAXIMALLY IN THE RED REGION OF THE SPECTRUM AND THE PFR FORM THAT ABSORBS MAXIMALLY IN THE FAR-RED REGION. PHOTOCOVERSION OF PR IN PFR INDUCES AN ARRAY OF MORPHOGENIC RESPONSES, WHEREAS PHOTOCOVERSION OF PFR TO PR CANCELS THE INDUCTION OF THOSE RESPONSES. PFR CONTROLS THE EXPRESSION OF A NUMBER OF NUCLEAR GENES INCLUDING THOSE ENCODING THE SMALL SUBUNIT OF RIBULOSE-BISPHOSPHATE CARBOXYLASE, CHLOROPHYLL A/B BINDING PROTEIN, PROTOCHLOROPHYLLIDE REDUCTASE, RNA, ETC. IT ALSO CONTROLS THE EXPRESSION OF ITS OWN GENE(S) IN A NEGATIVE FEEDBACK FASHION.
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- PM: CONTAINS ONE COVALENTLY LINKED TETRAPYRROLE CHROMOPHORE.
CC -1- SIMILARITY: BELONGS TO THE PHYTOCHROME FAMILY.
CC -1- SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 HISTIDINE KINASE DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; U56698; AAB67863.1; -.
DR InterPro: IPR003018; GAF.
DR InterPro: IPR003594; HATPase_C.
DR InterPro: IPR004359; HIS_KIN_s1g.
DR InterPro: IPR003661; HIS_kina.
DR InterPro: IPR001610; PAC.
DR InterPro: IPR000014; PAS.
DR InterPro: IPR001294; Phytochrome.
DR Pfam: PF02519; GAF; 1.
DR Pfam: PF00989; PAS; 2.
DR Pfam: PF00360; phytochrome; 1.
DR Pfam: PF00512; signal; 1.
DR PRINTS: PR01033; PHYTOCHROME.
DR SMART: SM00065; GAF; 1.
DR SMART: SM00387; HATPase_C; 1.
DR SMART: SM00388; HISKA; 1.
DR SMART: SM00086; PAC; 1.
DR SMART: SM00091; PAS; 2.
DR PROSITE: PS50109; HIS_KIN; 1.
DR PROSITE: PS50112; PAS; 2.
DR PROSITE: PS50245; PHYTOCHROME_1; 1.
DR PROSITE: PS50046; PHYTOCHROME_2; 1.
KW Transcription regulation; Photoreceptor; Phytochrome; Chromophore;
KW Repeat; Multigene family.
FT DOMAIN 608 679 PAS 1.
FT DOMAIN 742 813 PAS 2.
FT DOMAIN 893 1113 HISTIDINE KINASE.
FT BINDING 319 319 CHROMOPHORE (BY SIMILARITY).
SQ SEQUENCE 1121 AA; 124126 MW; FFA0DB62426EA32E CRC64;

Query Match 74.6%; Score 763.5; DB 1; Length 1121;
Best Local Similarity 68.3%; Pred. No. 1.9e-69;
Matches 140; Conservative 28; Mismatches 28; Indels 9; Gaps 1;

QY 1 KLAARAIKRLQSLPGDGLCDYVEVYQRLTGIDRVWVYFHEDDHGEVSEIRSL 60
DB 198 KLAARAIKRLQSLPGDGLCDYVEVYQRLTGIDRVWVYFHEDDHGEVSEIRSL 257
QY 61 EPIGLHYPATIDIPQARFLFKONRYRMICDCAFPVAVIYDKRLQSLGSLTRAPH 120
DB 258 EPIGLHYPATIDIPQARFLFKONRYRMICDCAFPVAVIYDKRLQSLGSLTRAPH 317
QY 121 GCHTYMANMGSVASIALAIYV--KGKDS-----KLMGLVGHGHSPPRYVPEP 171
DB 318 GCHTYMANMGSVASIALAIYV--KGKDS-----KLMGLVGHGHSPPRYVPEP 377
QY 172 CEFLMQAFGLQLOMELQSLAEK 196
DB 378 CEFLMQAFGLQLOMELQSLAEK 402

RESULT 12
PHYB_SOLTU STANDARD; PRT; 1129 AA.
AC P34094;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Phytochrome B.
GN PHYB.
OS Solanum tuberosum (Potato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
ON NCBI_TaxID=4113;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93081720; PubMed=1450376;
RA Heyer A., Gatz C.;
RT "Isolation and characterization of a cDNA-clone coding for potato type B phytochrome.";

RL Plant Mol. Biol. 20:589-600(1992).

CC -1- FUNCTION: REGULATOR PHOTORECEPTOR WHICH EXISTS IN TWO FORMS THAT

CC ARE REVERSIBLY INTERCONVERTIBLE BY LIGHT: THE PR FORM THAT ABSORBS

CC MAXIMALLY IN THE RED REGION OF THE SPECTRUM AND THE PER FORM THAT

CC ABSORBS MAXIMALLY IN THE FAR-RED REGION. PHOTOCONVERSION OF PR IN

CC PER INDUCES AN ARRAY OF MORPHOGENIC RESPONSES, WHEREAS

CC RECONVERSION OF PER TO PR CANCELS THE INDUCTION OF THOSE

CC RESPONSES. PER CONTROLS THE EXPRESSION OF A NUMBER OF NUCLEAR

CC GENES INCLUDING THOSE ENCODING THE SMALL SUBUNIT OF RIBULOSE-

CC BISPHOSPHATE CARBOXYLASE, CHLOROPHYLL A/B BINDING PROTEIN,

CC PHOTOCOLOROPHYLLIDE REDUCTASE, RNA, ETC. IT ALSO CONTROLS

CC THE EXPRESSION OF ITS OWN GENE(S) IN A NEGATIVE FEEDBACK FASHION.

CC -1- SUBUNIT: HOMODIMER.

CC -1- PTM: CONTAINS ONE COVALENTLY LINKED TETRAPYRROLE CHROMOPHORE.

CC -1- SIMILARITY: BELONGS TO THE PHYTOCHROME FAMILY.

CC -1- SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.

CC -1- SIMILARITY: CONTAINS 1 HISTIDINE KINASE DOMAIN.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@sib-sib.ch).

CC -----

CC EMBL: S51538; AAB24397.1; -.

CC PIR: S28431; S28431.

CC Mendel: 10718; SOLTU:PHYB.1.

CC InterPro: IPR003018; GAF.

CC InterPro: IPR003594; HATPase_C.

CC InterPro: IPR004359; HIS_KIN_S19.

CC InterPro: IPR003661; HIS_KIN_A.

CC InterPro: IPR000014; PAS.

CC InterPro: IPR001294; Phytochrome.

CC Pfam: PF01590; GAF. 1.

CC Pfam: PF02518; HATPase_C. 1.

CC Pfam: PF00989; PAS. 2.

CC Pfam: PF00360; phytochrome. 1.

CC Pfam: PF00512; signal. 1.

CC PRINTS: PRO1033; PHYTOCHROME.

CC SMART: SM00065; GAF. 1.

CC SMART: SM00387; HATPase_C. 1.

CC SMART: SM00388; HIS_KA. 1.

CC SMART: SM00091; PAS. 2.

CC PROSITE: PS50109; HIS_KIN. 1.

CC PROSITE: PS50112; PAS. 2.

CC PROSITE: PS00245; PHYTOCHROME_1; 1.

CC PROSITE: PS50046; PHYTOCHROME_2; 1.

CC KMW: Transcription regulation; Photoreceptor; Phytochrome; Chromophore;

CC Repeat: Multigene family.

CC FT DOMAIN: 622 693 PAS 1.

CC FT DOMAIN: 755 826 PAS 2.

CC FT DOMAIN: 903 1122 HISTIDINE KINASE.

CC BINDING: 333 333 CHROMOPHORE.

CC SEQUENCE: 1129 AA; 125621 MW; 60DF5FB964EFC60B CRC64;

QY Query Match 74.6%; Score 763.5; DB 1; Length 1129;

Db Best Local Similarity 74.7%; Pred. No. 1.9e-69;

Matches 142; Conservative 18; Mismatches 21; Indels 9; Gaps 1;

QY 16 GDIGALCDTVVEDYQRLTGYDRVMYGFHEDDHGEVSEIRKSDLEPLGLHYPATDIPQ 75

Db 227 GTLKLCCTVAVESYELGYDRVMYKFEHDEHGEVASEKRSKSDLEPLGLHYPATDIPQ 286

QY 76 AARFLFKONRMYTCCNATPYKVVQSEELKRLPLCLVNSTLAPRGCHTQYMANNGSVAS 135

Db 287 ASRFLFKONRMYTCCNATPYKVVQSEELKRLPLCLVNSTLAPRGCHTQYMANNGSVAS 346

QY 136 LALAIYVGRD-----SSKLMLGLVYGHHCGRYVPPFLRYACEFLMAQFGIQLQME 186

Db 347 LTLAVIINGNDEAVGGGRNSRMLGLVYGHHTSVRSIPFLRYACEFLMAQFGIQLQME 406

QY 187 IQLASQLAEK 196

Db 407 IQLASQLSEK 416

RESULT 13

ID PHY_ADICA STANDARD; PRT; 1118 AA.

AC P42496;

DT 01-NOV-1995 (Rel. 32, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 01-MAR-2002 (Rel. 41, Last annotation update)

DE Phytochrome 1.

DE PHY1.

OS Adiantum capillus-veneris (Fern).

OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;

OC Filicopsida; Filicales; Adiantaceae; Adiantum.

OX NCBI_TaxID=13818;

GN [1]

RP SEQUENCE FROM N.A.

RA Okamoto H., Hirano Y., Abe H., Tomizawa K.I., Furuya M., Wada M.;

RT "The deduced amino sequence of phytochrome B from seed plants.";

RT Plant Cell Physiol. 34:1329-1334(1993).

RM [2]

RP REVISIONS TO 28; 93-97; 310; 345 AND 1044.

RA Nozue K., Fukuda S., Kanegae T., Wada M.;

RL Submitted (JUL-1998) to the EMBL/Genbank/DBJ databases.

CC -1- FUNCTION: REGULATOR PHOTORECEPTOR WHICH EXISTS IN TWO FORMS THAT

CC ARE REVERSIBLY INTERCONVERTIBLE BY LIGHT: THE PR FORM THAT ABSORBS

CC MAXIMALLY IN THE RED REGION OF THE SPECTRUM AND THE PER FORM THAT

CC ABSORBS MAXIMALLY IN THE FAR-RED REGION. PHOTOCONVERSION OF PR IN

CC PER INDUCES AN ARRAY OF MORPHOGENIC RESPONSES, WHEREAS

CC RECONVERSION OF PER TO PR CANCELS THE INDUCTION OF THOSE

CC RESPONSES. PER CONTROLS THE EXPRESSION OF A NUMBER OF NUCLEAR

CC GENES INCLUDING THOSE ENCODING THE SMALL SUBUNIT OF RIBULOSE-

CC BISPHOSPHATE CARBOXYLASE, CHLOROPHYLL A/B BINDING PROTEIN,

CC PHOTOCOLOROPHYLLIDE REDUCTASE, RNA, ETC. IT ALSO CONTROLS

CC THE EXPRESSION OF ITS OWN GENE(S) IN A NEGATIVE FEEDBACK FASHION.

CC -1- SUBUNIT: HOMODIMER.

CC -1- PTM: CONTAINS ONE COVALENTLY LINKED TETRAPYRROLE CHROMOPHORE.

CC -1- SIMILARITY: BELONGS TO THE PHYTOCHROME FAMILY.

CC -1- SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.

CC -1- SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.

CC -1- SIMILARITY: CONTAINS 1 HISTIDINE KINASE DOMAIN.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@sib-sib.ch).

CC -----

CC EMBL: AB016168; BAA51856.1; -.

CC InterPro: IPR003018; GAF.

CC InterPro: IPR003594; HATPase_C.

CC InterPro: IPR004359; HIS_KIN_S19.

CC InterPro: IPR003661; HIS_KIN_A.

CC InterPro: IPR000014; PAS.

CC InterPro: IPR001610; PAC.

CC InterPro: IPR000014; PAS.

CC InterPro: IPR001294; Phytochrome.

CC Pfam: PF01590; GAF. 1.

CC Pfam: PF02518; HATPase_C. 1.

CC Pfam: PF00989; PAS. 2.

CC Pfam: PF00360; phytochrome. 1.

CC Pfam: PF00512; signal. 1.

CC PRINTS: PRO1033; PHYTOCHROME.

CC SMART: SM00065; GAF. 1.

CC SMART: SM00387; HATPase_C. 1.

CC SMART: SM00388; HIS_KA. 1.

CC SMART: SM00086; PAC. 1.


```

DT 01-JUN-1994 (rel. 29, Created)
DT 01-JUN-1994 (rel. 29, Last sequence update)
DT 01-MAR-2002 (rel. 41, Last annotation update)
DE Phytochrome 1.
GN PHY1.
OS Physcomitrella patens (Moss).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;
OC Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.
OX NCBI_TaxId=3218;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94039823; PubMed=8224238;
RA Kolukisaoglu H.U., Braun B., Martin W.F., Schneider-Poetsch H.A.W.;
RT "Mosses do express conventional, distantly B-type-related
RT phytochromes. Phytochrome of Physcomitrella patens (Hedw.).";
RL FEBS Lett. 334:95-100(1993).
CC -1- FUNCTION: REGULATORY PHOTORECEPTOR WHICH EXISTS IN TWO FORMS THAT
CC ARE REVERSIBLY INTERCONVERTIBLE BY LIGHT: THE PR FORM THAT ABSORBS
CC MAXIMALLY IN THE RED REGION OF THE SPECTRUM AND THE PFR FORM THAT
CC ABSORBS MAXIMALLY IN THE FAR-RED REGION. PHOTOCONVERSION OF PR IN
CC PFR INDUCES AN ARRAY OF MORPHOGENIC RESPONSES, WHEREAS
CC RECONVERSION OF PFR TO PR CANCELS THE INDUCTION OF THOSE
CC RESPONSES. PFR CONTROLS THE EXPRESSION OF A NUMBER OF NUCLEAR
CC GENES INCLUDING THOSE ENCODING THE SMALL SUBUNIT OF RIBULOSE-
CC BISPHOSPHATE CARBOXYLASE, CHLOROPHYLL A/B BINDING PROTEIN,
CC PROTOCHLOROPHYLLIDE REDUCTASE, RNA, ETC. IT ALSO CONTROLS
CC THE EXPRESSION OF ITS OWN GENE(S) IN A NEGATIVE FEEDBACK FASHION.
CC -1- SUBUNIT: HOMODIMER.
CC -1- PTM: CONTAINS ONE COVALENTLY LINKED TETRAPYRROLE CHROMOPHORE.
CC -1- SIMILARITY: BELONGS TO THE PHYTOCHROME FAMILY.
CC -1- SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 HISTIDINE KINASE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See: http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X75025; CAA52933.1; -.
DR PIR: S37206; S37206.
DR InterPro: IPR003018; GAF.
DR InterPro: IPR003594; HATPase_C.
DR InterPro: IPR004359; HIS_KIN_slg.
DR InterPro: IPR003661; HIS_KIN_A.
DR InterPro: IPR001610; PAC.
DR InterPro: IPR000014; PAS.
DR InterPro: IPR001294; Phytochrome.
DR Pfam: PF01590; GAF; 1.
DR Pfam: PF02518; HATPase_C; 1.
DR Pfam: PF00989; PAS; 2.
DR Pfam: PF00360; phytochrome; 1.
DR Pfam: PF00512; signal; 1.
DR PRINTS: PRO1033; PHYTOCHROME.
DR SMART: SM00065; GAF; 1.
DR SMART: SM00387; HATPase_C; 1.
DR SMART: SM00388; HISKA; 1.
DR SMART: SM00066; PAC; 1.
DR SMART: SM00091; PAS; 2.
DR PROSITE: PS50109; HIS_KIN; 1.
DR PROSITE: PS50112; PAS; 2.
DR PROSITE: PS00245; PHYTOCHROME_1; 1.
DR PROSITE: PS50046; PHYTOCHROME_2; 1.
KW Transcription regulation; Photoreceptor; Phytochrome; Chromophore;
KW Repeat.
FT DOMAIN 610 681 PAS 1.
FT DOMAIN 744 815 PAS 2.
FT DOMAIN 895 1115 HISTIDINE KINASE.
FT BINDING 321 321 CHROMOPHORE (BY SIMILARITY).
SQ SEQUENCE 1132 AA; 125230 MW; E1DAD4D6DC9CD16 CRC64;

```

```

Query Match          72.3%; Score 740.5; DB 1; Length 1132;
Best Local Similarity 66.3%; Pred. No. 4,1e-67;
Matches 136; Conservative 29; Mismatches 31; Indels 9; Gaps 1;

QY 1 KLAVRAISRLQSLPGDIGALCDTVEDYQRLTGDRAWYIQFHEDDHEVSEIRSDL 60
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 200 KLAAKAIFRLOALPGGNIGILCDTVVEEYRELGYDRAVMAVRFHEDEHEVAAEIRRADL 259
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 61 EPTLGLHYPATDIPQARFLFKONRYRMICDCAFPVKVQSEELKRPCLVNSTLRAPH 120
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 260 EPTLGLHYPTDIPQASRFLFKKNKRYITADCSAPPVKYIQPTLRQPVSLAGSTIRSPH 319
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 121 GCHTQYMANMGVASYALALAIYVKGDS-----KLGLVYGHCSPRVYPPPLRYA 171
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 320 GCHAQYMGNMGSIASLVMAVITINDNEDSHGSVQGRKLMGLVYCHHTSPRTVPPPLRSA 379
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 172 CEFLMQAFGLQLOMELQLASQLAEK 196
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 380 CGFLMQVFGLOLMEVESAAQLREK 404
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

Search completed: June 7, 2002, 18:58:04
Job time: 243 sec